DOCUMENT RESUME

ED 144 138 CS 501 831

TITLE	Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, April 1-June 30, 1977.
INSTITUTION	Haskins Labs., New Haven, Conn.
REPORT NO	SR-50 (1977)
PUB DATE	Jun 77
NOTE	253p.
EDRS PRICE	MF-$0.83 HC-$14.05 Plus Postage.
DESCRIPTORS	*Acoustic Phonetics; Articulation (Speech); *Artificial Speech; *Audition (Physiology); *Auditory Perception; Child Language; *Language Research; Phonemes; Phonetics; *Psychoacoustics

ABSTRACT

This report is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. The ten papers treat the following topics: speech synthesis as a tool for the study of speech production; the study of articulatory organization; phonetic perception; cardiac indices of infant speech perception; cinefluorographic and electromyographic studies of articulatory organization; stimulus dominance and ear dominance in the perception of dichotic voicing contrasts; voicing-conditioned durational differences in vowels and consonants in the speech of three- and four-year-old children; electromyography signal processing for speech research; perceptual test of a phonological rule; and adaptation of the category boundary between speech and nonspeech. A cumulative index to papers in previous collections and document numbers for previous collections are included. (AA)

****
* Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original. *
****

ED144138

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SR-50 (1977)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 April - 30 June 1977

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports).

CS 501 831

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

National Institute of Dental Research
Grant DE-01774

National Institute of Child Health and Human Development
Grant HD-01994

Assistant Chief Medical Director for Research and Development,
Research Center for Prosthetics, Veterans Administration
Contract V101(134)P-342

United States Army, Department of Defense
Contract MDA 904-77-C-0157

National Institutes of Child Health and Human Development
Contract N01-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-5596

National Science Foundation
Grant BNS76-82023

National Science Foundation
Grant MCS76-81034

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper, Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey, Treasurer
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Thomas Baer
Peter Bailey[1]
Fredericka Bell-Berti*
Gloria J. Borden*
Robert Crowder*
Donna Erickson
William Ewan
Eli Fischer-Jørgensen[3]
Carol A. Fowler*
Frances J. Freeman*
Jane H. Gaitenby
Thomas J. Gay*
Katherine S. Harris*
Alice Healy*
Isabelle Y. Liberman*
Leigh Lisker*
Ignatius G. Mattingly
Paul Mermelstein
Seiji Niimi[2]
Lawrence J. Raphael
Bruno H. Repp
Philip E. Rubin
Donald P. Shankweiler
Linda Shockey
George N. Sholes
Michael Studdert-Kennedy*
Quentin Summerfield[1]
Michael T. Turvey*
Robert Verbrugge*

Technical and Support Staff

Eric L. Andreasson
Elizabeth P. Clark
Harriet Greisser*
Donald Hailey
Terry Halwes
Elly Knight*
Sabina D. Koroluk
Agnes McKeon
Nancy R. O'Brien
Loretta J. Reiss
William P. Scully
Richard S. Sharkany
Leonard Szubowicz
Edward R. Wiley
David Zeichner

Students*

Mark Blechner
Steve Braddon
David Dechovitz
Laurel Dent
Susan Lea Donald
F. William Fischer
Hollis Fitch
Anne Fowler
Nieba Jones
Lynn Kerr
Morey J. Kitzman
Andrea G. Levitt
Roland Mandler
Leonard Mark
Nancy McGarr
Georgia Nigro
Mary Jo Osberger
Sandra Prindle
Abigail Reilly
Robert Remez
Helen Simon
Emily Tobey
Betty Tuller
Harold Tzeutschler
Michele Werfelman

---

*Part-time
[1] Visiting from The Queen's University of Belfast, Northern Ireland.
[2] Visiting from University of Tokyo, Japan.
[3] Visiting from Institute of Phonetics, University of Copenhagen, Denmark.

CONTENTS

I. Manuscripts and Extended Reports

Speech Synthesis as a Tool for the Study of Speech Production -- Franklin S. Cooper, Paul Mermelstein and Patrick W. Nye 1

The Study of Articulatory Organization: Some Negative Progress -- Katherine S. Harris 13

Phonetic Perception -- Alvin M. Liberman and Michael Studdert-Kennedy 21

Cardiac Indices of Infant Speech Perception: Orienting and Burst Discrimination -- Cynthia L. Miller, Philip A. Morse and Michael F. Dorman 61

Cinefluorographic and Electromyographic Studies of Articulatory Organization -- Thomas Gay 77

Stimulus Dominance and Ear Dominance in the Perception of Dichotic Voicing Contrasts -- Bruno H. Repp 93

Voicing-Conditioned Durational Differences in Vowels and Consonants in the Speech of Three- and Four-Year Old Children -- Lawrence J. Raphael, Michael F. Dorman and Donna Geffner 115

EMG Signal Processing for Speech Research -- Diane Kewley-Port . . . 123

Perceptual Test of a Phonological Rule -- Linda Shockey 147

Adaptation of the Category Boundary Between Speech and Non-speech: A Case Against Feature Detectors -- Robert E. Remez 151

II. Cumulative Index (SR-1 - SR-49) 171

- Author Index 172
- Title Index 211

III. Publications and Reports 237

IV. Appendix: DDC and ERIC numbers (SR-21/22 - SR-48) 239

I. MANUSCRIPTS AND EXTENDED REPORTS

Speech Synthesis as a Tool for the Study of Speech Production*

Franklin S. Cooper, Paul Mermelstein and Patrick W. Nye

ABSTRACT

Those elements of articulation that are essential to the communicative role of speech can be studied in ways that have proved successful in discovering the acoustic cues for speech perception. The method proposed is to synthesize speech with an appropriate articulatory model and evaluate the output signal by ear. In this way, hypotheses about articulatory cues for speech perception can be tested directly in terms of the intelligibility of the synthetic speech.

Control of the synthesis--by hand or by rule--will be in terms of movements of the model's articulators, but with access available also to intermediate stages for manipulation of vocal tract shape or acoustic spectrum. A primary consideration is to make the control of synthesis highly interactive; that is, displays and controls will be conceptually convenient and easy to operate, and the synthetic speech will be available for listening very soon after changes are made in the control parameters. Quality and naturalness of the synthetic speech are secondary considerations, since the main objective is good intelligibility with minimal articulatory descriptions.

Our intent in this paper is to describe some research studies that we are undertaking and to explain our reasons for choosing speech synthesis and the class of research questions that synthesis as a methodology implies. Briefly, we wish to learn what parts of the complex articulatory events of speech production are actually carrying the message, that is, what articulatory cues the speaker must produce in order that the listener will understand what was said. We think of this as a search for the articulatory cues that parallels earlier work we have done on searching for the acoustic cues in speech.

There are close parallels between the two kinds of search, and we have found it useful in planning the work on articulatory cues to draw analogies with our experience in searching for acoustic cues. Hence, we will speak of that experience in presenting our plans. We will even digress into a brief description of a new pattern playback we have built; it will be useful in the planned studies even though it was designed primarily for research in acoustic cues.

---

*This paper will appear in the U.S. Japan Joint Seminar on Dynamic Aspects of Speech Production, ed. by M. Sawashima and F. S. Cooper, Univ. of Tokyo Press.

[HASKINS LABORATORIES: Status Report on Speech Research SR-50 1977]

We have usually spoken of speech synthesis as a tool for the study of speech perception. But the acoustic cues we found all seemed to point back to articulation, implying that we were, in fact, studying production by way of perception. Thus, the parallels between our earlier work and the planned work can be viewed in this way: both were concerned with speech production, though the earlier work was on cues at the acoustic level, whereas the planned work is on cues at the articulatory level.

In either case, the distinguishing characteristics of the methodology are that it seeks to find the principal carriers of information, that it tests for these cues by perceptual methods, and that it uses synthetic speech to do so. Obviously, speech is the required stimulus when the perception of a message is to be tested, and synthetic speech has the very great advantage that systematic manipulation of the stimuli is possible, either at the acoustic level or at the articulatory level that precedes it.

Research Methods: from Acoustic Cues to Articulatory Cues

The method we used in searching for the acoustic cues, often called "hypothesize-and-test", proved well suited to that task (Liberman and Cooper, 1972). We think it will be equally effective in the search for the articulatory cues. The earlier work was, in fact, modeled on the chemist's customary technique of testing his analytic conclusions by synthesizing the suspected compound and comparing properties. We started with the patterns we thought we could see in sound spectrograms and regenerated sound from such patterns with a device we built for that purpose, namely, the Pattern Playback. In using it, a speaker produces an utterance from which the experimenter prepares a spectrogram. Guided by this spectrogram, a schematic copy is painted and passed to the Pattern Playback for conversion into synthetic speech. Now the two speech samples, the natural and the synthetic, are compared to determine by ear whether the essential acoustic cues have survived in the painted copy. The procedure is highly interactive. The user is given the opportunity to rapidly insert or delete spectral features at will and to immediately assess their importance by listening to the synthetic output and comparing it with the natural speech sample.

The principal ways in which we propose to model our new procedures on the old are by providing the means to obtain results quickly, to make modifications to the data interactively by hand, and to compare the outputs at a variety of different levels, but especially at the perceptual level. The organization of the research method is illustrated in Figure 1, which shows three ways to experiment on speech that is generated by a real speaker, an articulatory model, or a terminal analog speech synthesizer.

For articulatory synthesis, we may compare the articulatory control data for a particular articulation with EMG data from our physiological research, especially as to relative timing of events. Likewise, we may compare, almost directly, the vocal tract shape for articulatory synthesis with X-ray and fiberoptic data measured from an actual vocal tract. Differences in the moment-by-moment vocal tract configurations will indicate where improvements might be made in the synthesis. When rules have been used to compute the control signals and vocal tract configurations, means will also be available to override these controls and to make changes in the vocal tract shape

8

FIGURE 1

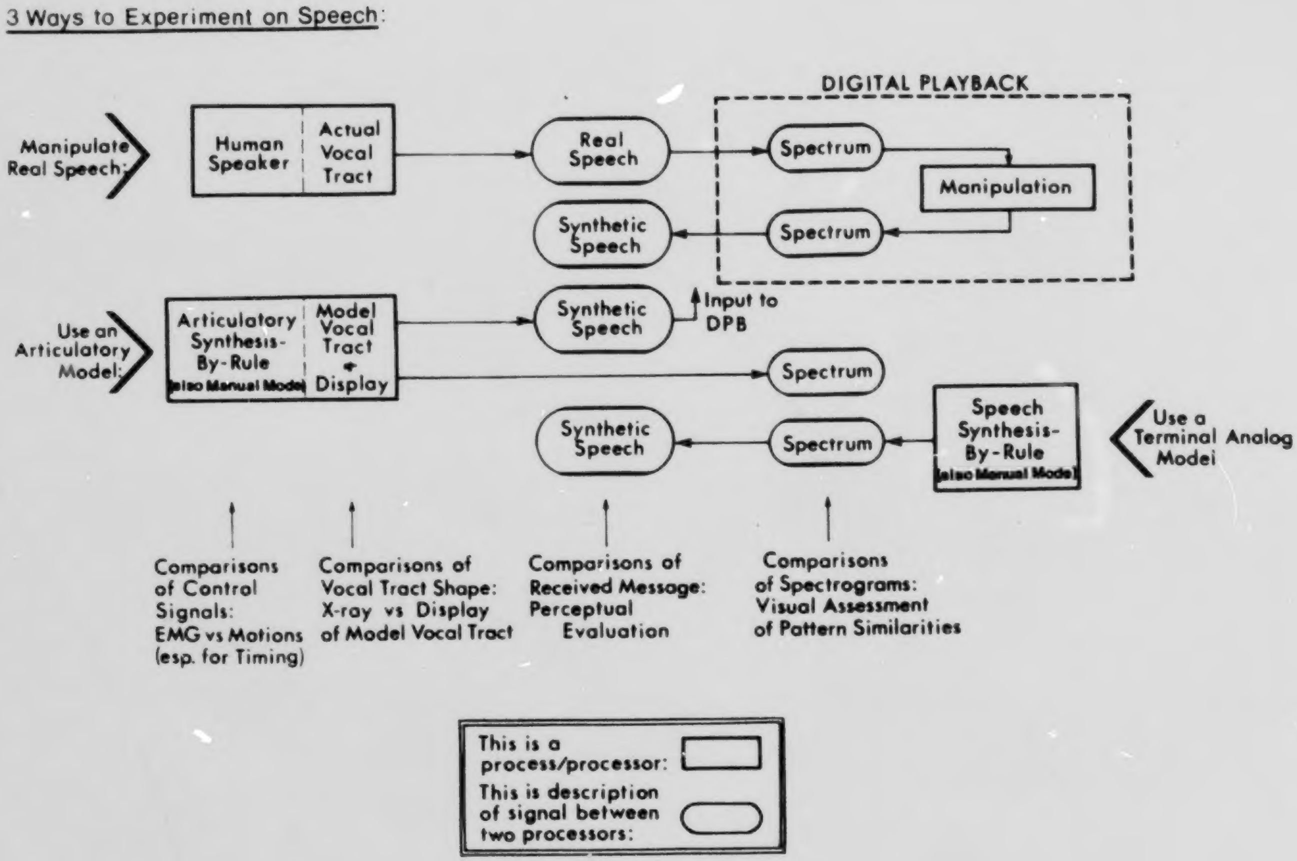

3 Ways to Experiment on Speech:
DIGITAL PLAYBACK
Manipulate Real Speech:
Human Speaker
Actual Vocal Tract
Real Speech
Spectrum
Manipulation
Spectrum
Synthetic Speech
Use an Articulatory Model:
Articulatory Synthesis-By-Rule
(also Manual Mode)
Model Vocal Tract + Display
Synthetic Speech
Input to DPB
Spectrum
Synthetic Speech
Spectrum
Speech Synthesis-By-Rule
(also Manual Mode)
Use a Terminal Analog Model
Comparisons of Control Signals: EMG vs Motions (esp. for Timing)
Comparisons of Vocal Tract Shape: X-ray vs Display of Model Vocal Tract
Comparisons of Received Message: Perceptual Evaluation
Comparisons of Spectrograms: Visual Assessment of Pattern Similarities
This is a process/processor:
This is description of signal between two processors:

Figure 1: Three ways to experiment on speech: (1) real speech can be manipulated by analysis-synthesis methods to discover the acoustic cues for speech perception; (2) an articulatory model can be used to generate synthetic speech and thereby test hypotheses about articulatory cues for speech perception; (3) likewise, a terminal analog synthesizer can be used to test hypotheses embedded in rules for synthesis. The key operation in all three kinds of experimentation is perceptual evaluation of the synthetic speech, coupled with visual comparison of the spectra of synthetic and real speech.

directly by hand. This facility will be useful in a number of experimental situations where it is desirable to examine the acoustic effects of individually specified articulatory movements.

As a final step in the above procedures, the output signal is presented to listeners, who are asked to make relative judgments about the speech, or absolute judgments about its intelligibility or adequacy. Exploratory manipulations and informal listenings will usually be followed by formal group tests.

Modeling the Speech Process

In representing the speech process by a model (or synthesizer) and in manipulating it with control parameters that specify the phonetic elements of the message, the choice of level of representation is crucial. Moreover, that choice hinges on a number of considerations: intended use, feasibility, conceptual convenience, and available knowledge are the primary desiderata.

If we consider human speech production, we find three distinct levels of the articulatory process that lie downstream from the presumed neural levels (to which we have little or no direct experimental access):

1. The activity of the individual muscles (in response to neuromotor commands).
2. The positions of the articulators and their movement in responses to muscle activity.
3. The corresponding vocal-tract shape in terms of the cross-sectional area function of the vocal tract.

For the research purposes we have in mind, namely an exploratory search for the articulatory cues, the third and lowest articulatory level is not very useful since, at the level of vocal-tract area functions, the conceptually important entities--the positions and movements of individual articulators--have already been merged into a single continuum. We will certainly wish to observe the performance of the model at this level, and even to exercise supervisory control over the area functions, but primary conceptualization and control can be done to better advantage at the next higher level, that is, by manipulating the articulators themselves.

Would we gain by working at the still higher levels of muscle activity or of the neuromotor commands that control the muscles? The philosophical question of where maximum simplicity is eventually to be found has yet to be convincingly answered. For the present, then, we rely on the practical considerations that our knowledge (from articulatory phonetics and cinefluorography) is better at level two than at level one, and that starting higher in the speech process would require more parameters and an additional stage of computation (to reach level two) without compensating advantages other than that electromyographic information could be applied more directly. For all these reasons, we intend to concentrate on the representation of phones and features at the second of the levels listed above and on transformations from that level to the speech signal.

Design Considerations for an Articulatory Synthesizer

We intend to start our research using an articulatory model developed by Mermelstein (1973) that allows parametric specification in the midsagittal plane for the position of the lips, tongue tip, tongue body, velum, jaw, and hyoid; we will extend the model by the addition of a variable that produces concave/convex arching of the tongue blade. These parameters permit the computation of vocal-tract transfer functions for laryngeal excitation or for fricative excitation at points internal to the tract (Mermelstein, 1972). The model has already demonstrated a capability for matching vocal tract configurations seen in X-ray movies and for generating highly intelligible VCV syllables.

The model does not simulate the entire speech-production system in man. In particular, it separates control of the sources of excitation of the vocal-tract resonances from control of the changes in those resonances with time. Since many aspects of coarticulation depend only on the interaction of the supraglottal articulators, only the positions of these articulators are computed, starting from a phonetic specification. Laryngeal excitation parameters (amplitude, fundamental frequency, onset, and duration) are specified explicitly, and effects of the supraglottal system back on the excitation source are neglected. Similarly, for frication, the amplitude and spectrum of the noise is explicitly specified; the output spectrum will, of course, reflect not only the source spectrum but also the filtering action of the vocal-tract cavities posterior and anterior to an assumed frication source at the point of maximum constriction along the tract.

The prime reason for not modeling directly the effects of articulator movement on the characteristics of the sound generation process is to limit the complexity of the simulation. Thus, we do not for the present intend to model laryngeal action because we do not think that it plays a central role in the coarticulation processes that we plan to study first. An exception may be the relative timing of laryngeal and supralaryngeal events, but this does not require detailed simulation of laryngeal mechanisms. Similarly, although the generation of frication is directly dependent on appropriate articulatory conditions, its accurate modeling requires very precise timing and positioning of the articulators. For these reasons, we rely on explicit control over the excitation signal itself rather than over the generative processes. The perceptual effects of simultaneous excitatory and articulatory variations can still be evaluated quite adequately, despite these substitutions for aerodynamic effects that link excitation to articulation in real speech.

The considerations that led to modeling articulatory movements exclusively in terms of the resulting vocal-tract shape in the midsagittal plane were primarily based on observational limitations. That is, the model was originally developed on the basis of a systematic examination of a series of midsaggital X-ray tracings of the vocal tract, in conjunction with the time-synchronized speech signal. By working interactively with the model during its development, it could be shown that displacements of the midsagittal vocal-tract outline can be derived from the movements of the independently controlled articulators.

Primary control of the model in terms of positions and movements of the principal articulators is, of course, an essential design consideration: this mode of control is conceptually convenient for the experimenter and is the natural framework for the application of structural and dynamic constraints. The articulatory model we will be using builds on the "ball-in-mouth" model of the articulators that was introduced by Coker and Fujimura (1966), but uses a more nearly complete set of articulators. The parameters assigned to these articulators are position variables that indicate the position of the structure in fixed space or relative to some other articulatory structure to which the articulator is primarily attached. For example, lip and tongue-body positions are specified with respect to the moving jaw. This representation allows an active mode of movement when an articulator's own parameters are changing; alternatively, a passive mode of movement may be executed relative to the fixed articulators as a result of movement of the structure to which that articulator is attached, but relative to which its position remains unchanged.

The model first computes the midsagittal outlines that result from the momentary positions of the articulators and then computes the midsagittal separations relative to an essentially fixed outer structure. Published information is used to convert these distances, measured at a large number of points along the vocal tract, to a continuous cross-sectional area function along a center-line distance function between the glottis and the lips. Up to 25 area samples spaced 0.875 cm apart are now computed and used in a nonuniform acoustic transmission line representation. Appropriate lumped terminations model the larynx at one end of the tract and the lips at the other. Articulations accompanied by a velar opening are modeled with the aid of an acoustic sidebranch that parallels the nonuniform transmission line for the oral tract. The cross-section of this nasal branch is assumed to be fixed except for a region near the velum.

The Control and Display of Articulatory Synthesis

The articulatory process will, of course, be simulated on a computer, since digital simulation provides flexibility and convenience that is not attainable through the use of physical models. For the model to be a truly useful tool, it must be equipped with displays that allow observation of the consequences of input instructions at all levels of the synthesis process. Further, to facilitate the hypothesize-and-test mode of experimentation, the model is controllable by interactive graphical editing at either the level of the individual articulators or of vocal-tract shape; also, when comparison with spoken utterances is desired, changes can be made directly in the spectral representation. Finally, the model must provide an acoustic output promptly on demand. Only on this basis can the user readily assess the perceptual consequences of the synthesis process, and only when the synthesizer responds promptly to changes in the control parameters is it easy to maintain a conceptual link between the hypothesis being tested and the result of the test.

The control and display facilities we will use are best considered in terms of the functional modes in which the model is to be operated. At the articulatory level, there is need for convenient control in terms of articulatory parameter values (for the individual articulators) and their allowed variation. As an aid in visualizing these numerical specifications, the corresponding vocal-tract outline (midsagittal) will be displayed. To change or improve the synthetic sound, interactive graphical editing will allow the user to redraw part or all of a midsagittal vocal-tract outline. X-ray information can be conveniently introduced at this point. Generally, also, the articulatory parameters may be quickly determined from such an outline.

When the model is used in its dynamic mode, the specified articulations for allophones will be supplemented by a set of rules that govern the time-variation of the articulatory system. With the specified articulations stored in a table of parameter values, one for each specified allophone, the rules will operate on the selected sets of these parameters to yield continuous functions of time.

An alternative procedure to the automatic generation of articulatory sequences by rule--one that will be especially useful in trying out new ideas or making detailed improvements to rule-generated sequences--is to redraw individual parameter "tracks" on the interactive graphic display. The total effect (on the other parameters as well) can then be seen in the midsaggital display and heard as the synthetic speech output.

At the spectrum level, it will be useful to view the spectral consequences of the articulatory movements. For voiced articulations, it is advantageous to view the spectral envelope without regard to fundamental-frequency variations. Such a spectral envelope and the corresponding formant frequencies can be derived from the model without the need for generating the actual signal waveform. Since the formant frequencies are the terms in which the acoustic cues are best known, this also makes for easy comparisons. When the results, as viewed in the above representations, are acceptable, we will generally want to generate the acoustic signal itself. This is so because, by listening to the signal, we may quickly judge its quality or naturalness and assess its identifiability.

At this point we can make good use of another research tool we are just completing: the Digital Pattern Playback (Nye, Reiss, Cooper, McGuire, Mermelstein and Montlick, 1975). This device stores the speech spectrum in computer core memory and so can immediately display a conventional gray-scale spectrogram for interactive graphical editing. Visual comparisons can then be made between the original spectrogram (generated, in this case, by articulatory synthesis) and either (1) the same spectrogram after it has been edited to improve intelligibility, or (2) a spectrogram of human speech of the same sentence. The Digital Pattern Playback also provides for comparison by ear of the sounds that correspond to the spectrograms. In other ways too, the DPP's capabilities for display and manipulation of speech data make it a useful companion device to the articulatory synthesizer.

Ideal Synthesizers and Research Synthesizers: Why They May Differ

Speech synthesis based on articulatory models has, of course, a considerable history. Some of the major contributions to the design and use of articulatory synthesizers, and to the underlying knowledge about relations between articulation and sound, are listed under References. There have been varied reasons for building such synthesizers; in some cases, the reason was to demonstrate that synthesis could be done in a particular way, in some, to mimic human production as seen by X-ray, and in some to attempt the production of more natural speech than is easily obtainable from terminal analog synthesizers or to control synthesis at a lower bit rate. Usually, some part of the effort has been directed to getting natural-sounding speech, that is, to approximating the performance of an ideal synthesizer.

It seems obvious that an articulatory synthesizer deserving the label "ideal" would have a capability for mimicking human speakers quite exactly. We would rate its performance initially on the naturalness of its "spoken" output; later, we would inquire about how accurately its chain of transformations (from phonetic sequence to sound) match those of the human speaker. Comparisons would be made at the levels of vocal-tract shape and acoustic spectrum--perhaps, even at the level of the speech waveform. It hardly needs saying that no existing synthesizer comes near to meeting such criteria.

However, ideal performance is not necessarily what we most want from a research synthesizer; that is, the question of ideal performance needs reexamination when we ask, not about the naturalness of the speech, but about the usefulness of the synthesizer for research--in particular for searching out the articulatory cues. The objective of this latter task is to find the simplest possible description of articulatory events that will, despite crudities of the speech, let a listener recover the phonetic message.

If we draw on our experience in searching for the acoustic cues, we will wish to manipulate these articulatory "events" in a variety of ways to study the perceptual consequences for the corresponding speech events. Sometimes this will involve efforts at simplification, for example, by allowing only the tongue or the lips to move in synthesizing a syllable that is normally spoken with some degree of movement by most of the articulators. Again, experimentation will involve stepwise variation in the relative timing of two component gestures, for example, of tongue and lip movements in synthesizing a syllable like [ibu] or an initial consonantal cluster such as [bℓ] in [bℓed]. Here good discrimination of the time delay between lip and tongue release would imply a basically unified organization for the cluster, whereas poor discriminability would indicate an independence of the constituent gestures. Too much delay of the tongue-lip release would result in the insertion of a vowel in the perceived phonetic string (that is, [bəℓed] instead of [bℓed]); of course, this must be avoided, since it would cue a phonetic distinction.

Obviously, experimental manipulations of this kind do not mimic natural speech. Often they call for an independence of control or a grading of spatial and temporal relationships that a human speaker could hardly achieve. To be sure, they ought not violate physiological constraints but, short of that, we will want to put the articulators through their paces in order to assess the perceptual consequences. Our expectation about the resulting sounds is that many, perhaps most, will sound "strange" but that some, perhaps many, will be clearly identifiable.

Thus, simplicity, both conceptual and operational, is a primary requirement in a research synthesizer. We expect to employ the fewest independent articulators, and the fewest control parameters to position and move them, that will still generate acceptable tokens for all the syllables of the language, that is, that will generate all the phones in the full range of phonologically allowed contexts.

Indeed, it is the essence of modeling to try for the maximum simplicity that will still give the required performance. In a search for the cues, performance is properly judged at the level of intelligibility, which is different from, and less demanding than, naturalness. Hence, naturalness in synthesis is for us not a primary short-term goal, nor was it needed in our earlier search for the acoustic cues. We found, in our work with the Pattern Playback, that the pursuit of the acoustic cues could proceed in the presence of a somewhat unnatural speech quality that even lacked pitch inflection. Intelligibility was the important requirement and proved to be nearly orthogonal to the dimension of naturalness.

Departures from naturalness are not, of course, a virtue, nor do we wish, when manipulating the articulators, to depart unnecessarily from the general configurations we see in X-ray movies. The guidance that level-by-level comparisons (of synthesis vs. nature) can give us is too valuable to be ignored. Indeed, we will sometimes want to manage the articulation so as to make it come quite close to the human model. The problem in designing a research synthesizer was to retain this capability, or as much as could be had, without paying too high a price in complexity of representation and control.

SUMMARY

Our reasons for undertaking a search for the articulatory cues are, first, that this will provide an insight into the nature of speech production comparable to the view we gained of speech perception when we succeeded in finding many of the major acoustic cues; and second, that the relationships between cues and phonetic elements should be simpler and more direct in the articulatory domain than they proved to be in the acoustic domain.

The origins of our interest in this undertaking lie in what we think we have learned about the nature of speech from two parallel lines of investigation. From studies of how speech is perceived, we learned that, although the acoustic signal contains a wealth of detail, only some of the things one sees in a spectrogram are important to the ear in identifying the phonetic content of the spoken message. These we have called the acoustic cues. Numerous other things that can be seen in the spectrogram are largely irrelevant, at least for intelligibility. By ignoring these things and synthesizing speech from patterns that contained only the acoustic cues, we were able to greatly simplify the acoustic signal and still retain most of the intelligibility. When we examined the nature of these acoustic cues, however, we found few one-to-one correspondences between them and the phones they represented; rather, the relationships were complex in ways that pointed to a reorganization and overlapping of articulatory gestures during speech production.

From physiological studies of how speech is produced, we have learned that articulatory events also seem complicated; thus, articulation, as seen in X-ray motion pictures or electromyographic recordings, involves most of the articulators most of the time. We can assume that here, too, some limited set of the component gestures provides the critical information (by way of sound as intermediary) on the basis of which a listener identifies the phonetic content of the message. These we refer to as the articulatory cues. If our assumption is correct, then much of the total articulatory description is also largely irrelevant, at least for intelligibility. But interest in the articulatory cues goes beyond stripping away irrelevancies. A more important point follows from our interpretation of the nature of their counterparts in the acoustic domain: if the acoustic cues do indeed reflect their articulatory origins, then the articulatory cues should show the simpler relationship with phonetic elements.

We think the time is right to undertake a search for the articulatory cues. There exists an extensive body of knowledge about both perception and production. There is a proven research method and experience with a computer-based articulatory synthesizer on which to implement it. Thus, both a significant problem and the means to probe it are at hand.

BIBLIOGRAPHY

Chiba, T. and N. Kajiyama. (1941) The Vowel, Its Nature and Structure. (Tokyo: Tokyo-Kaiser Kan Pub. Co.).

Coker, C. H. and O. Fujimura. (1966) Model for specification of the vocal tract area function. J. Acoust. Soc. Am. 40, 1271.

Coker, C. H., N. Umeda and C. P. Browman. (1973) Automatic synthesis from ordinary English text. IEEE Trans. Audio Electroacoust. AU-21, 293-298.

Dunn, H. K. (1950) The calculation of vowel resonances, and an electrical vocal tract. J. Acoust. Soc. Am. 22, 740-753.

Fant, G. (1960) Acoustic Theory of Speech Production. ('s-Gravenhage: Mouton).

Flanagan, J. L., K. Ishizaka, and K. L. Shipley. (1975) Synthesis of speech from a dynamic model of the vocal cords and vocal tract. Bell System Tech. J. 54, 485-506.

Fujimura, O. (1962) Analysis of nasal consonants J. Acoust. Soc. Am. 34, 1867-1875.

Fujimura O. (1975) Syllable as a unit of speech recognition. IEEE Trans. Acoust. Speech Sig. Proc. ASSP-23, 79-82.

Liberman, A. M. and F. S. Cooper. (1972) In search of the acoustic cues. In Papers in Linguistics and Phonetics to the Memory of Pierre Delattre, ed. by A. Valdman. (The Hague: Mouton), pp. 329-338.

Liberman, A. M., F. S. Cooper, D. P. Shankweiler and M. Studdert-Kennedy. (1967) Perception of the speech code. Psychol. Rev. 74, 431-461.

Lindblom, B. E. F. and J. Sundberg. (1971) Acoustical consequences of lip, tongue, jaw and larynx movements. J. Acoust. Soc. Am. 50, 1166-1179.

Mermelstein, P. (1971) Calculation of the vocal-tract transfer function for speech synthesis applications. In Proceedings of the Seventh International Congress on Acoustics, Vol. 3. (Budapest: Akademiai Kiado), pp. 173-176.

Mermelstein, P. (1972) Speech synthesis with the aid of a recursive filter approximating the transfer function of the vocal tract. In Conference Record, 1972 Conference on Speech Communication and Processing, April 24-26, Newton, Mass., pp. 152-155.

Mermelstein, P. (1973) Articulatory model for the study of speech production. J. Acoust. Soc. Am. 53, 1070-1082.

Nye, P. W., L. J. Reiss, F. S. Cooper, R. M. McGuire, P. Mermelstein, and

T. Montlick. (1975) A digital pattern playback for the analysis and manipulation of speech signals. Haskins Laboratories: Status Report on Speech Research SR-44, 95-107.

Öhman, S. (1966) Coarticulation in VCV utterances: spectrographic measurements. J. Acoust. Soc. Am. 39, 151-168.

Peterson, G. E. and H. L. Barney. (1952) Control methods used in a study of the vowels. J. Acoust. Soc. Am. 24, 175-184.

Rice, L. (1971) A new line analog speech synthesizer for the PDP-12. Working Papers in Phonetics 17 (Linguistics Department, University of California at Los Angeles), 58-75.

Rosen, G. (1958) Dynamic analog speech synthesizer. J. Acoust. Soc. Am. 30, 201-209.

Stevens, K. N. (1968) On the relations between speech movements and speech perception. Zeitschrift f. Phon. u. Kommunikationsforschung 21, 102-106.

Stevens, K. N. and A. S. House. (1955) Development of a quantitative description of vowel articulation. J. Acoust. Soc. Am. 27, 484-493.

Stevens, K. N. and A. S. House. (1963) Perturbation of vowel articulations by consonantal context. J. Speech Hearing Res. 6, 111-128.

The Study of Articulatory Organization: Some Negative Progress*

Katherine S. Harris†

ABSTRACT

This paper examines some of the evidence against some commonly-held views of the nature of articulatory units. Four hypotheses are examined: First, speech perception, and hence speech production, operates on some invariant "core unit" in a time-varying signal. Second, speech perception extracts invariant units from a time-varying signal on a phoneme target basis. Third, unit targets are supported by positioned feedback. Fourth, the difficulties with the first three formulations can be solved by changing from "phoneme" units to some higher level unit. Reasons are found for discarding all these hypotheses.

INTRODUCTION

This paper is an attempt to summarize what we now know, or rather don't know, about a vaguely defined area called "the organization of speech." In particular, the topic is what MacNeilage has called the "reality status of concepts of linguistic units." (MacNeilage, 1973).

The study of the organization of speech is the province of speech science, a rather uneasy blend of elements from phonetics and motor physiology. Perhaps I can illustrate the mixture with an anecdote quoted from Granit's (1967) biography of the great neurophysiologist, Sherrington. He is recorded as saying to his student, Wilder Penfield, "It must be nice to hear the preparation speak to you." Our hypotheses about speech organization came partly, then, from phonetics, and partly from general neurophysiology. I would like to run through four of these hypotheses, in their argument form, and discuss some evidence that has caused them to fail. The first hypothesis follows:

---

*This paper was presented at the U.S. Japan Joint Seminar on Dynamic Aspects of Speech Production and appears in the proceedings Dynamic Aspects of Speech Production: Current Results, Emerging Problems, and New Instrumentation, Hill-Top Hotel, Tokyo, 7-10 December, 1976.

†Also Graduate School and University Center, City University of New York.

Acknowledgement: This work has been supported in part by a grant from the National Institute of Dental Research.

HYPOTHESIS 1

> Speech is perceived as having invariant units: therefore, perception must operate on invariant parts of the acoustic signal.

This hypothesis, or versions of it, guided early work at the Bell Telephone Laboratories. It has two obvious problems. The first is that the acoustic signal for a given speech sound and for a given speaker, depends on the size of the vocal tract. This problem led Peterson (1952, 1961), and later Gerstman (1968) to suggest that the listener arrives at vowel judgments by some kind of perceptual normalization of the presented vowel, based on the relationship between formants. There are some problems with this theory as a dynamic hypothesis, as we will discuss below, but refinements of the theory will account for differences between steady-state formant values of the vowels for different speakers.

A second problem, and at that time an apparently more serious one, was that when convenient visual displays for the acoustic speech signal became widely known, no units corresponding to phonetic entities were obvious (Potter, Kopp and Green, 1947). A suggestion made by them was that perception is organized to focus on the relatively steady-state aspects of the signal, skipping over the variable "transitional" stretches between those steady states. Indeed, Cyril Harris (1953), then working at Bell, attempted to synthesize speech by putting together short segments of speech clipped from the ongoing stream. The result was unintelligible.

I don't believe we have yet learned quite enough from that failure. Even at the time, a different interpretation of the transitional portions was available, namely, that these transitions were essential for speech intelligibility, since they could be shown to have cue value, particularly for the consonants. This interpretation had, indeed, been demonstrated directly in work with speech synthesis (Cooper, Delattre, Liberman, Borst and Gerstman, 1952). There were later attempts to "explain" the speech perceptual mechanism by more complicated hypotheses, as we will see below.

HYPOTHESIS 2

> Speech is perceived in terms of invariant units: it can be shown that there are few steady-state segments in speech. Hence, speech perception must process the signal to extract invariant perceptual units from a time varying signal.

This is one version of the Haskins "motor theory," stripped of its physiological detail (Liberman, Cooper, Harris and MacNeilage, 1963). Basically, the idea is that, in production, there is temporal and spatial smear of various low-level aspects of the articulatory process, so that the resulting acoustic output is an encipherment of the input signal; further, in perception, the perceptual apparatus somehow decodes the signal, by reference to articulation, into its underlying units. There are a number of subhypotheses, of varying degrees of sophistication, about what these underlying units might be (Harris, 1976).

Invariant Electromyographic Signals

There were some early Haskins attempts to show that the signals to the muscles were less variable than the resulting acoustic outputs (Harris, Lysaught and Schvey, 1965; MacNeilage, 1963; Cooper, Liberman, Harris and Grubb, 1958; Cooper, 1965). Apart from the difficulty of testing the proposition that one type of unit is less variable than another, the hypothesis suffers from the fact that, as stated, it ignores the variations in muscle signal size associated with the different distances through which articulators must travel, when different phonetic units are juxtaposed. This point was discussed by MacNeilage (1970) who observed that coarticulation effects on muscle signals, due to this effect, are ubiquitous.

Articulatory Targets

The point of view that articulatory movement "aims at" articulatory targets, is the view espoused by MacNeilage in the paper cited above. He suggests that the targets are maintained by some form of feedback from the periphery, as does Abbs (1973).

Acoustic Targets

A variant of this view, advanced by Ladefoged (1967) and Lieberman (1973), among others, is that speakers aim at acoustic targets, which can be realized by different articulatory maneuvers, depending on context or speaker.

Closely related views have been developed for somewhat different ends by Lindblom (1963) and by Öhman (1967). Lindblom, in explaining vowel neutralization in rapid or destressed speech, suggested that invariant signals are sent to the articulators for a given phoneme target, but that the target is not always attained because the next signal may be sent too soon, causing target undershoot. Öhman (1967), in attempting to account for phonetic context effects, suggests that they arise from the temporal overlap of movements towards target positions. Lindblom has developed a very similar inertial view of speech timing effects (Lindblom, 1967) to account for differences in the inherent duration of vowels.

All these models have a common view of the speech process: peripheral encoding is believed to account for coarticulation of signals which are invariant at a central stage in the articulatory process. To the extent that these models specify a perceptual process, they assume, either explicitly or implicitly, that perception proceeds by reversing the encoding operations of production (Lindblom and Studdert-Kennedy, 1967).

Recent evidence suggests that this is a misleading picture. Strange, Verbrugge, Shankweiler and Edman (1976) presented listeners with sets of natural vowels, either alone or in consonantal context. They found that identification was better when the vowels were in context. If perception were indeed a steady-state target-extracting process of any kind, it would be hard to explain these results. When a listener is presented with vowels in steady-state form, they are presumably already "at target." When they are presented in CVC context, the vowel target must usually be inferred, due to undershoot or similar adjustments. Since no decoding is required in the former case, listeners should be maximally accurate. The fact that they perform worse calls into question any of the "target-extraction" formulations.

HYPOTHESIS 3

Speech has invariant units. These units are maintained in adult speakers by some form of nonacoustic feedback from the periphery.

Hypotheses about the role of feedback in speech have been with us for some time, although the literature has not always been explicit about what kind, or kinds, of feedback is crucial, as between gamma-loop feedback (Abbs, 1973), or tactile and kinesthetic feedback (Ringel and Steer, 1963). However, in spite of the general importance of the topic, there seem to be substantial roadblocks in the path of finding out more by the means presently available. Three approaches have been used.

First, there have been a number of studies involving reduction of oral tactile sensation, most notably those of Ringel and his associates (for example, Ringel and Steer, 1963). In general, these experiments show that the effects of block on various branches of the trigeminal nerve are not overwhelming (Scott and Ringel, 1971; Borden, Harris and Catena, 1973). Furthermore, the experimental procedure causes motor, as well as sensory effects, so that the results are difficult to interpret (Borden, Harris and Catena, 1973; Abbs, Folkins and Sivarajan, 1976).

A second approach has been to scan the relevant neurophysiological literature, in order to find an appropriate animal model for the human speech situation. While it is difficult for an amateur to assess the work, there does not seem to be an entirely appropriate analog for speech, and results are conflicting with regard to the importance of various types of feedback for various kinds of movement in the examples discussed. For example, animals can use deafferented limbs in learned or unlearned tasks (Taub and Berman, 1968), although some deterioration of fine motor control is generally found. On the other hand, lesions of the tract of the mesencephalic nucleus, which abolishes spindle afferent input from the masticatory muscles (Goodwin and Luschei, 1974) does not alter chewing behavior in any obvious way.

The third approach has been to study the effects of disruptions of articulation. Here again, there is no solid body of relevant experimentation and results are often conflicting. Folkins and Abbs (1975), for example, have shown that speakers can compensate immediately for the effects of unexpected interruptions of articulator movement. In their experiment, the jaw was unexpectedly loaded during the closure for a bilabial stop consonant. Results show that the lips compensate for the jaw in completing closure on the first trial. Another often cited study by Lindblom and Sundberg (1973) reports that a speaker can duplicate his natural vowels with a bite block between his teeth, with virtually no time for relearning. However, the only citation of the study I know is an oral report, with no experimental details. Hamlet and Stone (1976), using a different experimental paradigm, find compensatory effects over fairly substantial periods.

If articulation suffers some interference, the speaker may use either acoustic or nonacoustic feedback to compensate for the disruption. It has been suggested by Nooteboom (1970) that compensatory articulation may well be guided by acoustic rather than articulatory equivalence. If so, we would expect devastating effects of articulatory disruption accompanied by acoustic masking. So far as I know, this line of research is unexplored. Surely, the disruption experimental paradigm is eligible for far more searching exploration than it has thus far received.

HYPOTHESIS 4

Speech has units, but we would understand its organization better if we turned from phones to more appropriate units, such as:

Syllables

The evidence against Kozhevnikov and Chistovich's syllable based model of coarticulation (Kozhevnikov and Chistovich, 1965) is, in large part, a product of the industry of Kenneth Moll and his students (for example, Daniloff and Moll, 1968; McClean, 1973), although there has been recent interesting and important work by Benguerel (Benguerel and Cowan, 1974). These studies all show that there is little evidence that the syllable boundary, as traditionally defined, blocks coarticulation. Benguerel interprets his results as supporting a feature-based model of coarticulation, such as that of Henke (see below).

Features

This is not the place for an exposition of the virtues of feature-based models in general. However, some recent experiments argue against feature-based models of anticipatory coarticulation. Tom Gay, in his paper at this conference, will be discussing evidence that phonetic entities are separately organized at the electromyographic level, even when there is no reason for it in feature terms. Perhaps even more important is evidence that specifically contradicts Henke's "scan-ahead" model for articulatory coarticulation (1967). The model proposes that a given feature will appear in the speech stream as soon as it can, by assimilative spreading. Thus, if a nasal is preceded by a series of vowels which are unspecified for nasalization, they should all be equally nasalized. This does not happen (Kent, Carney and Sevareid, 1974; Ushijima and Hirose, 1974). The degree of nasality of a vowel, as measured by velar height during its production, depends on its proximity to a nasal consonant.

Some recent results of our own can be interpreted in the same way (Bell-Berti and Harris, 1976). We examined anticipatory coarticulation of lip rounding for /pɑsup/, /pɑtup/, /pɑtsup/ and /pɑstup/, measured by the electromyographic activity of the orbicularis oris muscle. We find that the onset of electromyographic activity seems to precede the onset of acoustic activity for the vowel by a fixed temporal interval, rather than to be locked to the preceding phone or cluster of phones.

Another bit of evidence arguing against a feature-based model of coarticulation comes from an experiment on Swedish rounded vowels (McAllister, Lubker and Carlson, 1974), again using the onset of orbicularis activity as a measure of anticipatory coarticulation of lip-rounding. They compared onsets of a series of front and back rounded vowels (both of which occur in Swedish) in the frame /itV/. If Henke's model were correct, lip rounding should begin at the same time, relative to the offset of /i/, for all vowels, since it is the feature composition of the preceding phones that determines the onset of anticipatory coarticulation. Interestingly enough, the onset of labial activity is later for the back vowels than for the front vowels, so that the lips seem to "wait for the tongue," which must move further for back vowels than for front vowels. In short, the temporal extent of anticipatory coarticulation cannot be predicted from a knowledge of the feature composition of the phones before the target.

One explanation of these data is that articulatory gestures are programmed temporally, and not in syllable or feature units; however, we have yet to determine the influence of stress and speaking rate on this timing. In addition, we must also examine the timing relationships between movements of different articulators, since we may find that subparts of segment gestures preserve their timing relationships.

Overall, given this rather negative review of our progress, what can we propose in a more positive direction? I can only offer a suggestion by my colleague, Michael Turvey (Turvey, Shaw and Mace, 1976), who points out in reviewing recent Russian studies of locomotion, that all skilled movements have subparts which tend to preserve their relationships to each other when the movement is transformed as by more rapid execution. He gives as an example the observation, by Kent, Carney and Sevareid (1974), that velar lowering and raising in the word contract is tied to particular events in the sequence of tongue movements. Whether this particular example suggests a useful experimental paradigm or not, it emphasizes a kind of observation we have been neglecting in studies of speech production: that is, what relationships between articulatory events are preserved when context changes, whether by increased speaking rate or stress, or segmental environment? Furthermore, how do the perceptual consequences of this view differ from those of a target extraction approach? Perhaps, when we can formulate experimental questions in terms such as these, we will be able to make progress in understanding speech organization.

REFERENCES

Abbs, J. H. (1973) The influence of the gamma motor system on jaw movements during speech: a theoretical framework and some preliminary observations. J. Speech Hearing Res. 16, 175-200.

Abbs, J. H., J. W. Folkins and M. Sivarajan. (1976) Motor impairment following blockade of the infraorbital nerve: Implications for the use of anesthetization techniques in speech research. J. Speech Hearing Res. 19, 19-35.

Bell-Berti, F. and K. S. Harris. (1976) EMG study of the coarticulation of lip rounding. J. Acoust. Soc. Am. 60, S63(A).

Benguerel, A.-P. and H. A. Cowan. (1974) Coarticulation of upper lip protrusion in French. Phonetica 30, 41-55.

Borden, G., K. S. Harris and L. Catena. (1973) Oral feedback II: An electromyographic study of speech under nerveblock anesthesia. J. Phonetics 1, 297-308.

Cooper, F. S. (1965) Research techniques and instrumentation: EMG. ASHA

Reports No. 1, 153-167.

Cooper, F. S., P. C. Delattre, A. M. Liberman, J. M. Borst and L. J. Gerstman. (1952) Some experiments on the perception of synthetic speech sounds. *J. Acoust. Soc. Am. 24*, 597-606.

Cooper, F. S., A. M. Liberman, K. S. Harris and P. M. Grubb. (1958) Some input-output relations observed in experiments on the perception of speech. Proceedings, 2nd International Congress on Cybernetics (Namur), 930-941.

Daniloff, R. and K. L. Moll. (1968) Coarticulation of lip-rounding. *J. Speech Hearing Res. 11*, 707-721.

Folkins, J. and J. H. Abbs. (1975) Lip and jaw motor control during speech: responses to resistive loading of the jaw. *J. Speech Hearing Res. 18*, 207-221.

Gerstman, L. J. (1968) Classification of self-normalized vowels. *IEEE Trans. Audio Electroacoust. AU-16*, 78-80.

Goodwin, G. M. and E. S. Luschei. (1974) Effects of destroying spindle afferents from jaw muscles on mastication in monkeys. *J. Neurophysiol. 37*, 967-981.

Granit, R. (1967) *Charles Scott Sherrington*. (Garden City: Doubleday).

Hamlet, S. L. and M. Stone. (1976) Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses. *J. Phonetics 4*, 199-218.

Harris, C. M. (1953) A study of the building blocks of speech. *J. Acoust. Soc. Am. 25*, 962-969.

Harris, K. S. (1976) Physiological aspects of speech production. *Haskins Laboratories Status Report on Speech Research SR-48*, 21-42.

Harris, K. S., G. F. Lysaught and M. M. Schvey. (1965) Some aspects of the production of oral and nasal labial stops. *Lang. Speech 8*, 135-147.

Henke, W. (1967) Preliminaries to speech synthesis based on an articulatory model. *Conference Preprints; 1967 Conference on Speech Communication and Processing*. (Bedford, Mass.: Air Force Cambridge Research Laboratories), 170-177.

Kent, R. D., P. J. Carney and L. R. Severeid. (1974) Velar movement and timing: Evaluation of a model of binary control. *J. Speech Hearing Res. 17*, 470-488.

Kozhevnikov, V. A. and L. A. Chistovich. (1965) *Rech', Artikulyatsiya, i Vospriyatiye*. Trans. as *Speech: Articulation and Perception*, 1966. (Washington, D.C.: Joint Publications Research Service), 30, 543.

Ladefoged, P. (1967) *Three Areas of Experimental Phonetics*. (London: Oxford University Press).

Liberman, A. M., F. S. Cooper, K. S. Harris and P. F. MacNeilage. (1963) A motor theory of speech perception. *Proc. Stockholm Speech Comm. Seminar, Vol. II*. (Stockholm: Royal Institute of Technology).

Lieberman, P. (1973) On the evolution of language: A unified view. *Cognition 2*, 59-94.

Lindblom, B. E. F. (1963) Spectrographic study of vowel reduction. *J. Acoust. Soc. Am. 35*, 1773-1781.

Lindblom, B. E. F. (1967) Vowel duration and a model of lip mandible coordination. Quarterly Progress and Status Report (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm) *STL-QPSR 4/1967*, 1-29.

Lindblom, B. E. F. and M. Studdert-Kennedy. (1967) On the role of formant transitions in vowel recognition. *J. Acoust. Soc. Am. 42*, 830-843.

Lindblom, B. E. F. and J. Sundberg. (1973) Neurophysiological representation of speech sounds. Paper presented at the XV World Congress of Logopedics and Phoniatrics, Buenos Aires, Argentina.

MacNeilage, P. F. (1963) Electromyographic and acoustic study of the production of certain clusters. J. Acoust. Soc. Am. 35, 461-468.

MacNeilage, P. F. (1970) Motor control of serial ordering of speech. Psychol. Rev. 77, 182-196.

MacNeilage, P. (1973) Linguistic units and speech production theory. J. Acoust. Soc. Am. 54, 329(A).

McAllister, R., J. Lubker and J. Carlson. (1974) An EMG study of some characteristics of Swedish rounded vowels. J. Phonetics 2, 267-278.

McClean, M. (1973) Forward coarticulation of velar movement at marked junctural boundaries. J. Speech Hearing Res. 16, 286-296.

Nooteboom, S. F. (1970) The target theory of speech production. IPO Annual Progress Report 5, Institute for Perception Research. (Eindhoven, Holland) 5, 51-55.

Öhman, S. E. G. (1967) Numerical model of coarticulation. J. Acoust. Soc. Am. 41, 310-320.

Peterson, G. E. (1961) Parameters of vowel quality. J. Speech Hearing Res. 4, 10-29.

Peterson, G. E. and H. L. Barney. (1952) Control methods used in a study of the vowels. J. Acoust. Soc. Am. 24, 175-184.

Potter, R. K., G. A. Kopp and H. C. Green. (1947) Visible Speech. (New York: Van Nostrand).

Ringel, R. L. and M. D. Steer. (1963) Some effects of tactile and auditory alterations on speech output. J. Speech Hearing Res. 6, 369-378.

Scott, C. M. and R. L. Ringel. (1971) Articulation without oral sensory control. J. Speech Hearing Res. 14, 804-818.

Strange, W., R. R. Verbrugge, D. P. Shankweiler and T. R. Edman. (1976) Consonant environment specifies vowel identity. J. Acoust. Soc. Am. 60, 213-224.

Taub, E. and A. J. Berman. (1968) Movement and learning in the absence of sensory feedback. In The Neuropsychology of Spatially Oriented Behavior, ed. by S. J. Freeman. (Homewood, Ill.: Dorsey).

Turvey, M. T., R. E. Shaw and W. Mace. (1976) Issues in the theory of action. Haskins Laboratories Status Report on Speech Research SR-48, 1-20.

Ushijima, T. and H. Hirose. (1974) Electromyographic study of the velum during speech. J. Phonetics 2, 315-326.

Phonetic Perception*

Alvin M. Liberman† and Michael Studdert-Kennedy††

INTRODUCTION

To include a chapter on phonetic perception in a handbook like this is to assume that the process is not wholly accounted for by such principles as we might find in research on the perception of nonspeech sounds. It is appropriate, then, that we here offer support for that assumption. We will not examine all relevant considerations, only those that bear most directly on the relation between the information in the acoustic signal and the listener's perceptual response to it; in our view, those are the most pertinent. Nor will we analyze such arguments as there are for the opposite assumption--namely, that auditory mechanisms are sufficient--though we will, as is proper, refer the reader to relevant papers.[1]

Phonetic perception is what happens when, on hearing speech, a listener recovers the phonetic message. That message consists of the meaningless segments we perceive as consonants and vowels. These are ordered in strings, organized into larger units, and carried on a prosodic contour. The segments, both consonants and vowels, are called 'phones'; among the larger units are syllables; the relevant aspects of prosody are stress and intonation. We must distinguish between the perceived phones and the more abstract phonologic forms that underlie them. Thus, the final segments in 'cats' and 'dogs' are different phones--voiceless [s] in 'cats' and voiced [z] in 'dogs'--yet at a

---

*To appear in Handbook of Sensory Physiology, Vol. VIII, "Perception," ed. by R. Held, H. Leibowitz, and H.-L. Teuber. Heidelberg: Springer-Verlag, Inc., 1977.

†Also University of Connecticut and Yale University.

††Also Queens College and Graduate Center of the City University of New York.

Acknowledgment: The preparation of this chapter was supported by a grant from the National Institute of Child Health and Human Development. Having freely borrowed the ideas of our colleagues at Haskins Laboratories, we are unable to say exactly which are owed to whom. We do especially acknowledge the contributions of Quentin Summerfield and Peter Bailey, who were guests at the Laboratories while this chapter was being prepared, and we are indebted to Mark Liberman for valuable suggestions.

[1] For reviews of speech perception research from several points of view, see Darwin, 1976; Pisoni, in press; Stevens and House, 1972.

more abstract phonologic (or morphophonemic) level they are the same. Our concern will be with the less abstract phones and their relation to the still less abstract sounds. Also, to keep our task within bounds, we will deal only with the segmental aspects of phonetic structure, including the organization of phones into syllables, though perception of prosody presents interesting, perhaps even similar, problems.

Students of language commonly assume a complex, grammatical relation between meaning and its phonetic vehicle, but often disregard the further complications that arise in the conversion to sound. They tend rather to suppose that the phonetic segments (or their constituent features) are represented discretely in the signal, as if by an acoustic alphabet. If that were so, perceiving phones would be like perceiving any other sounds; there would be no special problem of phonetic perception and no reason for this chapter. There is evidence, however, that the sounds of speech are not an alphabet on the phones, but a complex and grammatical code. In the first section of the paper we will place that code in the larger scheme of language and identify its important characteristics.

If it is true that the phones are linked to the sounds by a special code, we should suppose that extracting the phones from the sounds would require a correspondingly special decoder. In the second section we will give reasons for supposing that such a decoder may exist.

There is, of course, an alternative to grappling with the problems created by the peculiar relation between sound and phonetic message: we can try to evade them. Indeed, we might suppose that phonetic perception does not occur, that the segments of the phonetic level are mere fictions, invented by linguists for their convenience, with no functional significance in language or in its psychophysiology. On that view the listener would go directly from sound to some meaningful segment (for example, word), bypassing the phonetic and phonologic structure entirely. To justify our concern with phonetic perception, we will, in the third section, argue that phonetic (and phonologic) structure plays an important role in language and is, in fact, recovered by the listener when he comprehends what is said to him.

THE SPECIAL NATURE OF THE SPEECH CODE: FUNCTION, FORM, AND KEY

For anyone who would understand the perception of speech, the salient fact is that the perceived phones are related to the sounds by a peculiar grammatical code, one of several that link sound to meaning. To grasp the nature of that code, it is useful to view it as part of the larger grammar. (See, for example, Mattingly and Liberman, 1969; Liberman, 1970). For that purpose, we will divide grammar--and language--in two. Making the cut at the phonetic level, we will look first toward meaning and then, in the other direction, toward sound.

The Function of the Meaningless Phones and of the Grammatical Codes that Link Them to the Meaningful Message

To see what grammar accomplishes, and thus to appreciate the role of the meaningless phones, we should first consider the shortcomings of an agrammatic mode of communication. (See Liberman, Mattingly, Turvey, 1972; Liberman, in press). In that mode, there would be a straightforward connection between message and signal. Instead of grammatical rules like those that build longer and more complex structures (syllables and sentences, for example) out of shorter and simpler ones (phones and words), there would be only a list of all possible messages and their corresponding signals. Obviously, such a mode would work well enough if there were reasonable agreement in number between messages and signals. But, just as obviously, there is no such agreement: the number of messages we have to send is vastly greater than the number of holistically different signals we can efficiently produce and perceive, especially if we are committed to signaling with sound. In short, an agrammatic mode of communication would limit the number of possible messages to the small number of distinctively different sounds we can produce and perceive. The consequence would be that most of what we want to express with language would be inexpressible.

We should suppose, then, that one function of grammatical codes is to restructure the information in the messages so as to make it compatible with our sound-signaling ability, and thus to match the potentialities of the message-generating intellect to the limitations of the vocal tract and the ear. But why two grammars, syntax and phonology, and why the two kinds of segments, meaningful and meaningless, they govern? What is the function of this dual structure, characteristic of all languages, and especially of the meaningless, phonologic portion that concerns us in this chapter? Why not, in a simpler world, have only a syntax--rules that organize and reorganize segments (words, for example) that are meaningful? Such a language could, from a logical point of view, evade the limitations imposed by the paucity of different segments, since it would be possible, even with a small set, to construct an infinitude of messages. A phonology-free language would, of course, have to make do with a small vocabulary, but that is not, in logic, a devastating limitation: For to the extent that we can organize our semantic space by a hierarchy of features, a small vocabulary might nevertheless suffice for many of the things we want to talk about (Ogden, 1967). But specifying a particular thing would, at best, take a lot of talking and listening, given the properties of vocal tracts and ears, and it would require, in addition, that one's mind work in ways that may be uncongenial to it.

At all events, no language does get along with a very small vocabulary. Vocabularies tend to be large and to grow ever larger. (But see Klima, 1975, pp. 247-270). To achieve these large vocabularies, given the limited number of signals we can command, languages use a very few meaningless segments--two to three dozen, in most cases--to construct a large number of meaningful ones. Hence, phonology. Taken together, then, syntax and phonology serve as a kind of interface, joining an intellect, which initiates, comprehends, and stores messages, to a vocal tract and ear, which produce and receive the sounds by which those messages are conveyed (Mattingly, 1972; Liberman, 1974).

The Function of the (Grammatical) Speech Code that Links Phonetic Message to Sound

Perhaps the need for grammatical recoding has ended with the production of the phonetic message. If so, the final link to speech could be agrammatic--a unit of sound for each segment of the message--and thus of no special interest to either the linguist or psychologist. But the phonetic message is only a stage in the grammatical process that connects meaning to sound. Further and still quite drastic restructuring is necessary. To see why, we need only pit the most obvious requirements of phonetic communication against the capabilities of the ear and the vocal tract. Although that has been done in earlier publications (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Liberman, et al., 1972; Studdert-Kennedy, in press), we ought nevertheless to offer a brief review here.

Two requirements of phonetic communication are of special interest: the phones must be communicated at a high rate, and their order must be properly apprehended by the listener. With regard to rate, it is obvious enough that language is more efficient the more rapidly it is communicated. It is only slightly less obvious that language is hard to understand when it is communicated too slowly. Slow communication can create difficulties because the meaning of the longer segments is distributed in complicated ways among the shorter segments they comprise. Hence full comprehension of a sentence, for example, must wait on completion of a structure that is formed by the words. The requirement about order follows from the use of a small number of phonetic segments. If we are to keep the number of segments per word within bounds, we must respect order: a word like 'dam' must be distinguished from its mirror image, 'mad.'

It is plain that if the phonetic segments were transmitted agrammatically--that is, each phone by a discrete segment of sound--the requirements of phonetic communication would not be met. We could neither speak nor listen as fast as we need to--and, indeed, do--nor could the listener keep the segments in their proper order. Speaking rates vary considerably, but they reach 20 or 25 phones per second, at least for short stretches. Presumably, it would be impossible to speak that rapidly if, as in an agrammatic mode, the gestures were made discretely, one for every phone and each in its turn. And even if the speaker could articulate that fast, the listener could not resolve the sound segments that would result; at 20 or 25 acoustic segments per second, the units of sound (hence phones) would merge to produce, in perception, a buzz or pitch. Moreover, the listener would have difficulty identifying the order of such discrete sound units, even at rates low enough to permit him to resolve them. Given the results of research on nonspeech sounds (Warren, 1969; 1976b), we should suppose that he could distinguish permutations of segments, but only on the basis of overall differences in the perceived pattern, not by assigning each segment to its own place in a sequence. Surely, then, the grammatical restructuring that makes communication distinctively linguistic cannot end with the production of the phonetic message. At least one more grammatical conversion is necessary if the message is to be transmitted and perceived efficiently.

The Form of the (Grammatical) Speech Code that Links Phonetic Message to Sound; the Fit of Form to Function

In the conversion of abstract phones to concrete sounds there is a restructuring of information, designed as if to match the requirements of phonetic communication to the properties of the vocal tract and the ear. Though much that is important about this conversion remains to be learned, enough is known to enable us to see some of its important characteristics. Thus, we know that the segments are first broken down into something like the well-known articulatory features of place of production, manner of production, and voicing[2]. (For an explication, see, for example, Ladefoged, 1971). As speech is produced, those separate features are assigned to the appropriate and more-or-less independent parts of the articulatory apparatus; the component gestures made by those parts are organized into preplanned coding units longer than a phonetic segment: and the organized complex of gestures, representing features of each of several successive phonetic segments, is produced simultaneously or with considerable overlap. The result is that the coding unit--roughly a syllable in many cases--comprises segments whose component gestures (features) are thoroughly interleaved (Cooper, Delattre, Liberman, Borst, and Gerstman, 1952; Fant, 1962; Liberman et al., 1967; Cooper, 1972; Stevens and House, 1972; Studdert-Kennedy, 1975a). We will call that arrangement by its common name, coarticulation.

Coarticulation enables a speaker to produce phonetic segments at rates considerably higher than the rates at which he must change the states of his articulatory muscles (Cooper, 1972). Thus, he speaks faster than he could if each phonetic segment were represented by a unit gesture, produced in its proper turn as one of a sequence of gestures. But coarticulation has consequences for perception as well, enabling the listener to evade just those limitations of the auditory system we referred to earlier. Consider, again, that if the phonetic message were transmitted agrammatically--that is, one acoustic segment for each phonetic segment--then the temporal resolving power of the ear would make it impossible to perceive speech at the rates that we do, in fact, commonly attain. But, as we have seen, the relation between phonetic message and sound is not agrammatic in that sense. Rather, coarticulation effectively folds information about several successive phonetic segments into a single stretch of sound. Moreover, the overlapped activity of several different articulators--for example, lips and tongue--will often affect the same parameter of the sound--for example, the second formant.[3] At

---

[2] In the case of the consonants, place of production refers to where in the mouth--lips, alveolar ridge, or velum, for example--the consonant constriction is made; manner of production refers to an articulatory maneuver--velum closed or open, tract totally closed or only partly closed, for example--that is characteristic of phones with the same place of production; voicing distinguishes classes of phones having the same place and manner according to the state of the vocal cords--open or closed--at the beginning of the gesture.

[3] A formant is a peak in the resonance curve of the vocal tract. The center value of this peak, specified in Hz, is called the formant frequency.

any chosen instant of time, therefore, each acoustic parameter is (commonly) carrying information about more than one phonetic segment. [For fuller discussion, see Liberman, et al., (1967).] That being so, the limit on rate of phonetic perception caused by the temporal resolving power of the ear is no longer set by the number of phonetic segments transmitted per unit time, but by the considerably smaller number of acoustic segments into which those phonetic segments have been encoded. Just how much saving is effected in this manner depends, of course, on the size of the encoding unit; and that will surely vary according to the nature of the contiguous phones in the string, rate of articulation, and other factors that we only dimly understand. But a significant amount of encoding will almost always occur--most obviously within the syllable--and it will, at every rate of articulation, effectively reduce the number of discrete acoustic segments that must be perceived.

Consider, now again, the difficulty the auditory system would have in identifying the order of phonetic segments at even moderate rates of speech if each phonetic segment were represented by an acoustic segment. But inasmuch as the phonetic segments are not so represented, the problem of identifying order of discrete acoustic segments does not arise (Day, 1970; Liberman, et al., 1972; Cole and Scott, 1974; Dorman, Cutting, Raphael, 1975). Recall how successive phonetic segments are encoded into the same stretch of sound, and imagine, for example, simple cases like [ba] and [ab]. If these syllables are produced at moderately rapid rates of articulation, it will be true of both acoustic patterns that information about the consonant and the vowel is carried simultaneously from the beginning of the sound to its end. But given that the articulatory gestures have opposite directions in the two cases--from closed (consonant) to open (vowel) for [ba] and from open (vowel) to closed (consonant) in [ab]--the acoustic shapes of the two acoustic syllables will be different. Indeed, they will be mirror images: for [ba] the formants will be rising throughout; for [ab] they will be falling. Thus, information about the order of phonetic segments is present in the acoustic signal, not as discrete events in ordered sequence, but as variations in shape or form (Liberman, 1976).

The Key to the Speech Code

Suppose the speech code were entirely arbitrary. In that case, a perceiving device could only match the signal against a dictionary of auditory templates, just as if it were using a code book. Of course, the templates could not correspond to segments the size of phones but would, rather, have to be at least as large as the coding unit that encompasses the acoustic consequences of coarticulation. As we remarked earlier, we do not know exactly how large that unit is or how stable it might be in the face of variations in speaking rate, word and phrasal stress, and other conditions of articulation. We can only suppose that, at the smallest, the unit would have to be of approximately syllabic size, since there is normally so much coarticulation within syllable boundaries.

But the speech code is not arbitrary; there is a key that unlocks it. To see the nature of the key, and how it makes sense of the relation between message and signal, we need only remind ourselves that the peculiarities of the speech code are just those that are introduced by the speaker as he lends himself to the processes by which the message is encoded in the sound. When those processes are understood, their consequences can hardly appear arbitrary. Thus, the key to the code is in the manner of its production. We should remark parenthetically that in this respect, speech is like the rest of language and different from most other processes: all the complications of language that the hearer must cope with are only those that, as speaker, he 'knows' how to introduce; the complications of nonlinguistic perception, on the other hand, are typically not owing to the hearer (or viewer) but are, rather, external to him. At all events, the processes by which speech is produced make it possible to understand the relation between acoustic signal and phonetic message, however peculiar that relation might be.

Although knowing how speech is produced enables us to see why the complications of the code should be peculiar in the way they are, it does not provide an automatic decoding procedure. Thus, we now understand enough about the speech code to be able to synthesize speech by rule (Ingemann, 1957; Liberman, Ingemann, Lisker, Delattre, and Cooper, 1959; Kelly and Gerstman, 1961; Kelly and Lochbaum, 1962; Cooper, 1962; for a summary, see Mattingly, 1974). That is, we can build a mechanism that accepts as input a string of phonetic symbols and then, as output, delivers speech. Using rules for the conversion that can be either acoustic or articulatory, the synthesizer produces speech that is imperfect--reflecting our imperfect command of the code--but rather highly intelligible, nevertheless, and reasonably acceptable. Now if we could simply turn those rules around, we should have a working model for speech perception. Unfortunately, the rules for synthesis, like all grammatical rules, work in only one direction, downhill; they take us from message to signal but not the other way. Perhaps there are rules that go in either direction, but they have not yet been found. Thus, to suggest that a listener might use the rules as a key, is only to imply some kind of connection between perception and production, of which more later; the underlying mechanism is, at present, unknown.

THE SPECIAL PROCESSES OF PHONETIC PERCEPTION

Surely the most parsimonious way to account for phonetic perception is to invoke only those mechanisms that are more or less common to mammalian (or primate) auditory systems. (See Miller, in press). Can we suppose, then, that such processes are sufficient, or must we look to specializations of various kinds? If specializations do exist, are they in the form of auditory devices that are tuned to respond to the phonetically relevant parts of the speech signal? Or are they more accurately characterized as integral parts of a system, more linguistic than auditory, that is specialized to deal with the peculiarities of grammatical codes? In this section we will consider whether both such specializations might exist, the one to deal with the purely acoustic characteristics of the perceptually important parts of the signal, the other to cope with the grammatical code that relates the signal to the phonetic information it conveys.

Auditory Specializations for Extracting the Phonetically Relevant Information from the Speech Signal

Many important attributes of the speech signal, including some that carry a heavy load of phonetic information, are not physically salient. For example, although most of the linguistically important information is contained in the lowest three formants, the acoustic energy is not tightly concentrated there but is, rather, smeared diffusely over the entire speech spectrum. Or again, despite the fact that consonants carry a far heavier load of segmental phonetic information than do vowels, they are signaled by far less acoustically prominent portions of the spoken syllable. Thus, formant frequency shifts (transitions) that carry important, even essential, information about consonantal place of articulation often make excursions of hundreds of cycles in some 30 or 40 msecs. Since humans seem to have no difficulty in extracting that information, one is led to wonder whether there may not be devices in the auditory system specialized for that purpose. These devices would be analogous, perhaps, to the feature detectors found in other species.

We have in mind the example of the cat, in which Whitfield and Evans (1965) found single cells ("miaow" cells) responsive to the rate and direction of frequency change. Whitfield (1965) pointed out the possible relevance of this finding to the perception of formant transitions in speech when he suggested that such units might be "...a final link in the mechanism...by which speech-like and similar signals are processed" (p. 247). If Whitfield is correct, we would have, not an auditory specialization for language, but rather a general auditory device (perhaps typical of mammals) that is exploited by humans for linguistic purposes.

In fact, an auditory mechanism specialized for language may be difficult to demonstrate, since we obviously cannot apply to humans the electrophysiological techniques that have been used on animals. It may, however, be possible to approach the matter indirectly, as, for example by extending to speech the adaptation procedures originally developed in studies of vision. The first to do this were Eimas and Corbit (1973). With synthetic syllables (for example, [ba] vs. [pa]) that ranged along an acoustic continuum, these investigators used the techniques of adaptation to produce shifts in the position of the perceptual boundary. The results led them to speculate that their procedures had affected a pair of binary phonetic feature detectors, and that adaptation or fatigue of one detector functionally sensitized its opponent. Subsequent work (see Cooper, 1975, and Ades, 1976, for reviews) demonstrated analogous effects for other consonantal feature oppositions. If these effects were truly on phonetic features, they would only provide additional evidence for the 'reality' of such entities and offer still another method, though potentially a most useful one, for defining their boundaries.

More relevant to our concerns here, therefore, are adaptation studies like the one by Bailey (1973), which showed that the effect decreased with a decrease in spectral overlap between adapting and test syllables. This suggests that if feature analyzing systems were indeed being isolated, the features were auditory rather than phonetic. (For a relevant discussion, see Ades, in press.) The finding by Bailey assumes considerable importance from our point of view, because there is apparently no other kind of evidence for the existence of feature analyzing systems of an auditory sort. Unfortunately, the matter appears not to be that simple. Further investigation has shown that the degree of adaptation is contingent on so many other aspects of the synthetic continuum, including intensity (Ganong, 1975) and fundamental frequency (Ades, in press), that one may, in the end, be led to doubt the feature interpretation altogether. Perhaps, then, the achievement of the work on selective adaptation will have been to demonstrate the operation of distinct perceptual channels rather than the existence of feature detectors as such. Nevertheless, the investigators may have found a method for exposing processes that respond to linguistically significant parts of the speech signal, and thus to have made possible the discovery of auditory specializations for language.

Linguistic Specialization for Recovering the Phonetic Message

Even if auditory detectors of the kind just discussed do exist, they could do no more than extract from the acoustic signal those features that are phonetically relevant. They might thus solve problems created by the fact that speech is, in certain respects, a poor signal, but it would presumably remain to some other device, more phonetic than auditory, to deal with the different fact that speech is a special code. As we were at pains to point out earlier, the peculiar characteristics of the code arise from the way speech is produced, in particular, from coarticulation. We should suppose, then, that the distinguishing characteristic of the phonetic device would be that it somehow makes use of that circumstance (Cooper et al., 1952; Liberman, Delattre and Cooper, 1952). For the present, the emphasis should be on the word "somehow"; we do not wish to speculate about the underlying mechanism, if only because we cannot offer relevant data. But if there is a device that behaves, by whatever means, as if it 'understood' how speech is produced, then we should expect to find evidence for a link between perception and production. Indeed, it would be just such a linkage that would clearly characterize phonetic as against auditory perception (Studdert-Kennedy, 1976; Liberman and Pisoni, in press).

In the sections that follow, we will identify several kinds of support for the assumption that there is a phonetic perceiving device and, correspondingly, a phonetic mode of perception. Some of that support is indirect in that it depends on our inability to account for certain phenomena of speech perception in terms of what we now know of how the ear works and what it commonly does; but some is more direct, being based on putative differences between auditory and phonetic perception and, in some cases, on the apparent links to production that characterize the phonetic mode.

Coping with the segmentation. If there were an acoustic criterion that could directly divide the speech stream into segments corresponding in size to the phones, then we should see no need to invoke other-than-auditory processes. No matter how complex in structure the acoustic segments might be, we should suppose that correspondingly complex auditory processes would be equal to the job. As we have seen, however, one of the characteristics of the speech code is that the phonetic information is distributed in curious ways through the sound. This is the most striking disparity between acoustic signal and phonetic message and, from the standpoint of a perceiving device, the most troublesome. Indeed, the disparity is greater than our characterization of the speech code might have implied, since the sound segments do not map onto the phones either in the way they divide or in the way they group. Thus, rapid switches in sound source during the articulation of successive phones may spread the information about a single message segment through several acoustic segments (Fant, 1962, 1968), as when stop-consonant closure and release into a following vowel yield a brief silence, an explosive release, a period of aspirated noise, and a more-or-less abrupt voice onset. On the other hand, coarticulation may, as we have previously noted, cause the information about several message segments to be collapsed into a single segment of sound.

The severity of the problem is evidenced by the fact that it has resisted solution for many years, as much by those concerned with speech synthesis (Coker et al., 1973) as by those working on automatic speech recognition (Ainsworth, 1976). Both groups have been driven to acknowledge that segments the size of phones are not to be found as segments in the acoustic stream; the irreducible acoustic unit is of approximately syllabic dimensions, just as we would expect given the very earliest result of research with synthetic speech (Cooper et al., 1952; Liberman et al., 1952). In the first attempt to 'synthesize' speech by commuting (and concatenating) segments of sound excised from prerecorded utterances, Harris (1953) found that the 'building blocks' had to be larger than phones. Other investigators (Peterson, Wang, and Sivertsen, 1958) later reported some success in producing speech by concatenating prerecorded segments, but the segments they required were a numerous and varied assortment of syllables and 'phoneme dyads'. Significant improvements in this method of synthesis have recently been made by Fujimura (1975, in press), though again the unit must be larger than the phone. And now, even in synthesis by rule, Mattingly (1976) has found it advantageous to preorganize the phonetic segments into syllables and then use those larger units as input to his synthesis program. As for the work on automatic speech recognition, it has long been plain that segmentation into phones by a straightforward acoustic criterion is hardly possible (Hyde, 1972), though segmentation into syllables can be done reasonably well (Mermelstein, 1975).

The foregoing considerations and facts suggest that the phones are not directly given in perception but must rather be derived from a running analysis of the signal over stretches of at least syllable length. There is ample experimental evidence that this is so.

Consider, for example, the matter of segment duration and its role in the perception of phones. It is known that, in English, the contrast between unreleased voiced and voiceless stops in syllable-final position (for example, [ab] vs. [ap]) can be determined by the duration of the preceding vowel (Denes, 1955; Raphael, 1972). But what happens, then, if that vowel is itself preceded by the consonant-vowel transitions appropriate to, say, [b], as in [bab] vs. [bap]? Does the listener pay attention only to the duration of the preceding vowel? Presumably he cannot do that if, as we have suggested, the transition cues for the consonant simultaneously carry information about the vowel. And, indeed, he does not. According to recent experiments (Raphael, Dorman, and Liberman, 1975), the duration used by the listener to determine voicing in the final segment includes all, or almost all, of the transition cues for the consonant in the initial segment.

Given only that result, we might suppose nevertheless that the listener takes one part of the acoustic signal as consonant and another part as vowel, provided we further suppose that the voicing of a syllable-final stop is determined by the sum of the durations of consonant plus vowel. At least two other experiments suggest, however, that the listener does not compute consonant and vowel durations on different parts of the syllable. In one of these experiments[4], listeners were asked to adjust the duration of a steady-state vowel to match the duration of the medial vowel in a stop-vowel-stop syllable whose formants had parabolic trajectories. As determined by that simple and direct technique, the perceived duration of the medial vowel was found to include a significant portion of the consonant-vowel transitions.

The other experiment dealt with duration as a cue for the perceived identity of a medial vowel and, simultaneously, with the voicing of a final stop, for example, [bɛt], [bæt], [bɛd], [bæd].[5] The results clearly imply that the listener did not assign one part of the syllable duration to the vowel and another part to the consonant. Rather, it was as if he used the whole duration of the syllable, but used it twice: once to determine the identity of the vowel and again to determine whether the syllable-final stop was voiced or voiceless.

That the information about the phonetic segments is smeared through the syllable is indicated also by evidence that the flanking transitions in a CVC syllable are used to judge the identity of the medial vowel. For example, Ochiai and Fujimura (1971) recorded natural, but distinctly articulated words and observed no errors of vowel identification. However, when they presented 50 msec portions gated from the vowel centers, listeners' judgments frequently shifted in directions that could be explained by contextual assimilation. Even more striking are the results of Strange, Verbrugge, and Shankweiler (1976). They recorded nine vowels spoken in isolation, and the same nine vowels spoken in various CVC frames. Despite the increased acoustic complexity introduced by a dynamic syllable structure, listeners correctly identified the vowels significantly more often when they were presented in a consonantal frame, even a variable one, than when they were presented in isolation. Thus, for the purpose of identifying the vowels, the perceiving system used those parts of the syllable that also contained information about the consonants. That is yet another reflection of the complex relation in segmentation between signal and message. But it also shows that though the perceptual target is a vowel, for which static formant frequencies are often assumed (Peterson and Barney, 1952), the perceptual system nevertheless prefers the dynamic configuration of a syllable, perhaps because it can then take advantage of the many constraints inherent in the way the vocal apparatus works when it coarticulates.

---

[4] Mermelstein, Liberman and Fowler: personal communication.

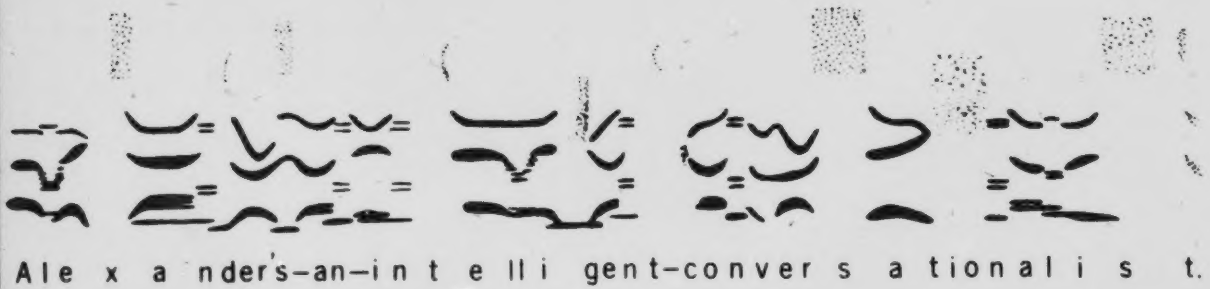
[5] Mermelstein: personal communication

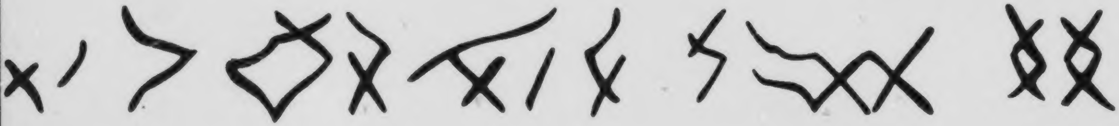
Considering all that is known about the peculiar disparity in segmentation between perceived message and transmitted signal, we suppose that the appropriately segmented percept lies at some remove from the immediately given auditory pattern, and that it is recovered by processes different from those the auditory system is ordinarily called on to provide. As for the possibility that such special processes make reference to production, we can offer no direct evidence about segmentation as such, only the observation that to find the segments, it must help to understand where they were lost.

Phonetic interpretation of the sounds of speech. We should now look more directly at some phenomena of speech perception that depend, presumably, on the same decoding processes that perform the segmentation but pertain more closely to what those segments, once retrieved, sound like to a listener. Do they sound like other sounds or do they not? And when not, is there evidence of a link to production?

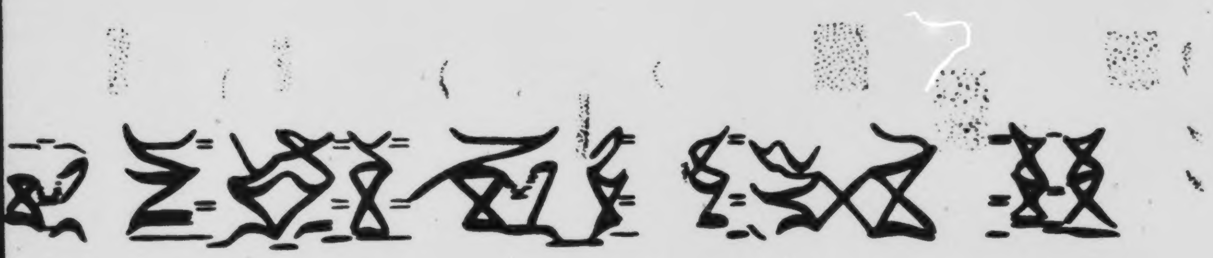
Impressions of the difference between auditory and phonetic modes. To convey a feeling for what we mean by the suggestion that the sound of speech is different from the sound of nonspeech, it may be useful to describe several phenomena that are part of the experience of people who work with synthetic speech. One of these is reflected in an observation made by investigators who used the Pattern Playback, an early research synthesizer that converted hand-painted spectrograms and other designs into sound (Cooper, 1950; Cooper, Liberman, and Borst, 1951; Cooper, 1953). Having succeeded in constructing highly schematized spectrograms, like the one at the top of Figure 1, that nevertheless produced intelligible speech, the investigators thought to take advantage of the flexibility of the Playback in order to destroy the intelligibility of speech by a novel and, they assumed, uniquely effective procedure: instead of drowning the speech in noise, which was the usual way, they would 'mislead' the ear. To that end they added to the spectrogram 'false' formants, always continuous with the 'true' formants, that improperly connected and extended the proper components of the acoustic pattern. An example is shown in the middle and at the bottom of Figure 1. In fact, as the reader can see, the eye is misled. But the ear was not. When the altered pattern was converted to sound, the listener heard the original phonetic message against a loud background of variously pitched whistles. It was as if the perceptual machinery had separated the acoustic effects that a vocal tract can produce from those it cannot. At all events, the effect was of two qualitatively different kinds of perception--articulate, monotone speech in the one case, complex and very bad 'music' in the other.

Much the same kind of phenomenon, though on a smaller scale, can be produced, not only on a device like the Pattern Playback, but also on the more modern parallel-resonance synthesizers now in common use. An example is seen in the contrast between the initial stop consonants of the syllables [ba] and [ga]. As shown in Figure 2, a sufficient acoustic cue is the direction of the second-formant transitions, rising for [b] and falling for [g]. Now, given our knowledge of psychoacoustics, we should suppose that those cues would sound like rising and falling glissandos or like chirps of different pitch, depending on how rapidly the formants moved on the frequency scale. And, in fact, when we present the formant-transition cues by themselves, as shown in the inset of Figure 2, that is exactly how they do sound (Mattingly, Liberman,


Alexander's-an-intelligent-conversationalist.



'Masking' Pattern



Composite

Figure 1: At the top, a hand-drawn spectrogram appropriate for synthesis of a sentence; in the middle, a pattern intended to 'mask' the sentence by 'misleading' the ear; and, at the bottom, the composite of sentence and 'mask,' which produces, again, the perception of the sentence, plus a dissociated set of whistles and noises.

Syrdal, and Halwes, 1971; Shattuck and Klatt, 1976). But what do we say, then, about the fact that those same transitions are heard in the context of speech as the abstract linguistic events we can only describe as [b] and [g]? Of course, the transition cues are isolated in the one case but part of a larger, if otherwise constant, pattern in the other, so that we might attribute the difference in perception to some kind of auditory interaction.

But even when the transition cues are in exactly the same acoustic context it is possible to hear them, simultaneously, as phonetic stops and auditory chirps. That effect was created by Rand (1974) in the following way. Into one ear he put all of the first formant and the steady-state parts of the second and third formants, while into the other ear he put just the transition cues (of the second and third formants) that distinguish [ba] and [ga], being careful to synchronize them properly with respect to the rest of the pattern. Though there is but one context--and indeed one brain--the formant transitions will, in this situation, often simultaneously produce two very different perceptions: the syllable [ba] (or [ga]) and a rising (or falling) chirp.

Essentially the same kind of effect has been created, though successively now instead of simultaneously, as part of an experiment designed by Bailey, Dorman, and Summerfield (1977) to permit a comparison of speech and nonspeech perception. The stimulus patterns are similar to those commonly used in research with synthetic speech in that they contain transitions appropriate to several stop consonant-vowel combinations, followed by vowel steady states; they differ from those normally used in that the formants are replaced by pure tones, one for each formant and set to its center of energy. On first being presented with such patterns, listeners hear them as a complex of tones, but after some time they begin to hear them as speech. We will not here presume to report on the results of the experimental comparisons that the study was designed to permit; we only remark the phenomenon, which is that there is a striking difference in subjective impression, depending on whether the listener is perceiving the stimulus patterns as tones or as speech; thus, it offers yet another way to gain a general appreciation of the perceptual differences between speech and nonspeech.

At all events, it is just such qualitative contrasts in perception as we have described here that can convey to a listener a direct impression of what we mean by the distinction between auditory and phonetic modes. We turn now to some relevant experimental observations.

Acoustic cues as a source of information about the speaker's vocal tract. Those aspects of the speech signal that, when varied, cause phonetically significant changes in perception are known as "acoustic cues." It is to those cues that we should now look, because we find there the clearest evidence for the link between perception and production that characterizes perception in the phonetic mode. No single piece of evidence is, by itself, wholly convincing; it is only the pattern that tells. For when we view the data in the light of known or imaginable auditory processes, we see a number of unconnected facts that require, apparently, an equal number of ad hoc assumptions. If we apply Occam's razor, however, we find a more or less comfortable fit to the single assumption that underlies this chapter: that the acoustic cues are processed, not only in the auditory system, but also at

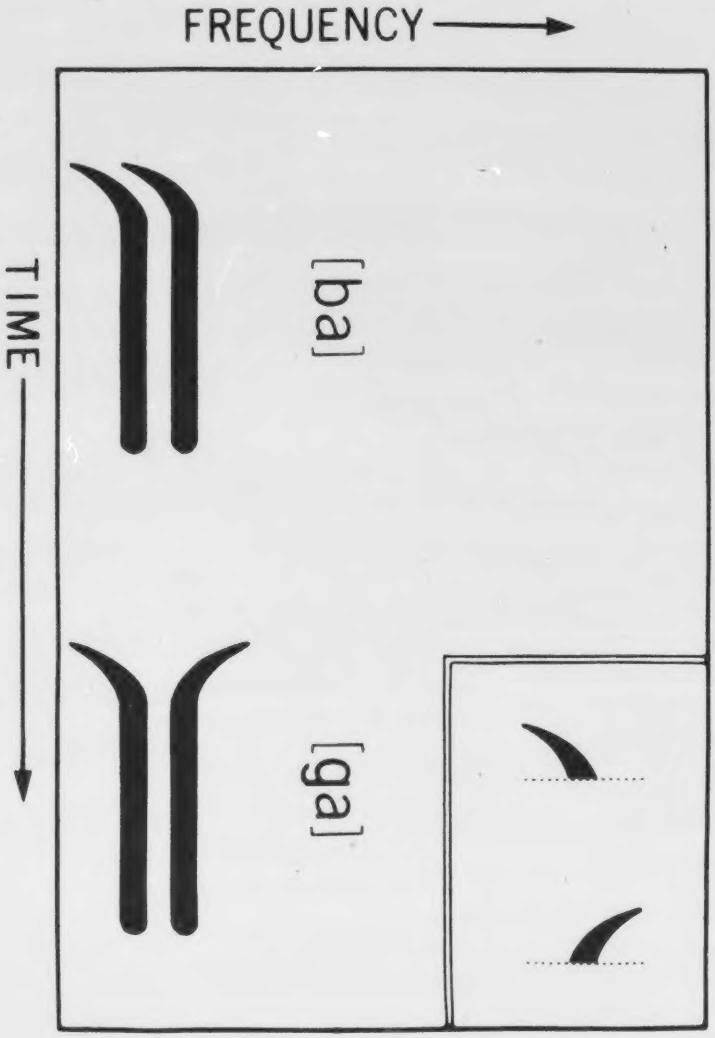


Figure 2: Spectrographic patterns sufficient for synthesis of [ba] and [ga]. Inset: The second-formant transitions that cue the perceived difference between the syllables, but sound, in isolation, like chirps.

some more abstract, phonetic remove; there, an appropriately specialized device uses the articulatory information provided by those cues to shape the listener's perception of what the speaker said.

A simple example. To see how an acoustic cue--silence--might provide information about a phonetically important gesture, we should consider the following facts about fricatives and stop consonants. A speaker cannot produce a stop consonant without closing his vocal tract for a brief period, and he cannot close his vocal tract without producing a period of silence. Hence, silence might be important to the perception of stop consonants, especially if the perceptual processes 'know' that stops require closure and that closure results in silence. It is relevant, then, to discover that in the perception of stops, silence is, in fact, an important condition.

Suppose, for example, we record the fricative-vowel syllable [sa]. As shown schematically in Figure 3, the acoustic pattern consists of a patch of noise, associated with the fricative, followed by a vocalic section. The vocalic section begins with the formant transitions characteristic of the fricative [s] when coarticulated with the vowel [a]; there follow, then, the steady-state formants characteristic of the (drawn-out) vowel [a]. It should be noted about the formant transitions at the beginning of the vocalic section that they are also appropriate, at least approximately, for the stop consonants [t] and [d], which have the same place of production as [s]. Now if we remove the patch of noise, listeners will commonly hear [ta], not [a]--that is, they will hear a stop consonant where none was before. If we now replace the s-noise in such a way as to create a silence of 50 or so msec between it and the vocalic portion, listeners will again hear the stop, this time in [sta]. We should say, parenthetically, that the same kind of effect can be obtained starting with a stop-vowel syllable like [ta]. In that case, putting s-noise immediately in front of the syllable will cause the listener to hear [sa], not [sta]; if the listener is to hear [sta], we must create a short period of silence between the s-noise and the vocalic section.

We see in this example that silence has just the sound we should expect it to have, given the assumption that it tells the listener whether or not the speaker closed his vocal tract long enough to have produced a stop consonant. But, surely, there might be other, perhaps more parsimonious, assumptions. We note in this connection that our examples conform to the paradigm for auditory forward masking, so we should take account of the possibility that the transition cues are simply being masked when the noise is too close to them; or we suppose, more vaguely, that there is some (not previously discovered) auditory interaction between silence and the transition cues which causes us to hear the peculiar sound of a stop consonant.

But there is considerable evidence that such alternative assumptions will not hold. Note, first, that in fricative-vowel syllables like the [sa] of our example, it has been found that the formant transitions contribute significantly to the perception of the fricative (Harris, 1958; Darwin, 1971). We should suppose, therefore, that the transition cues are 'getting through'--that is, they are not being masked by the s-noise. It is only their (phonetic) interpretation as fricative (when the silence is relatively short) or stop (when the silence is relatively long) that is affected.

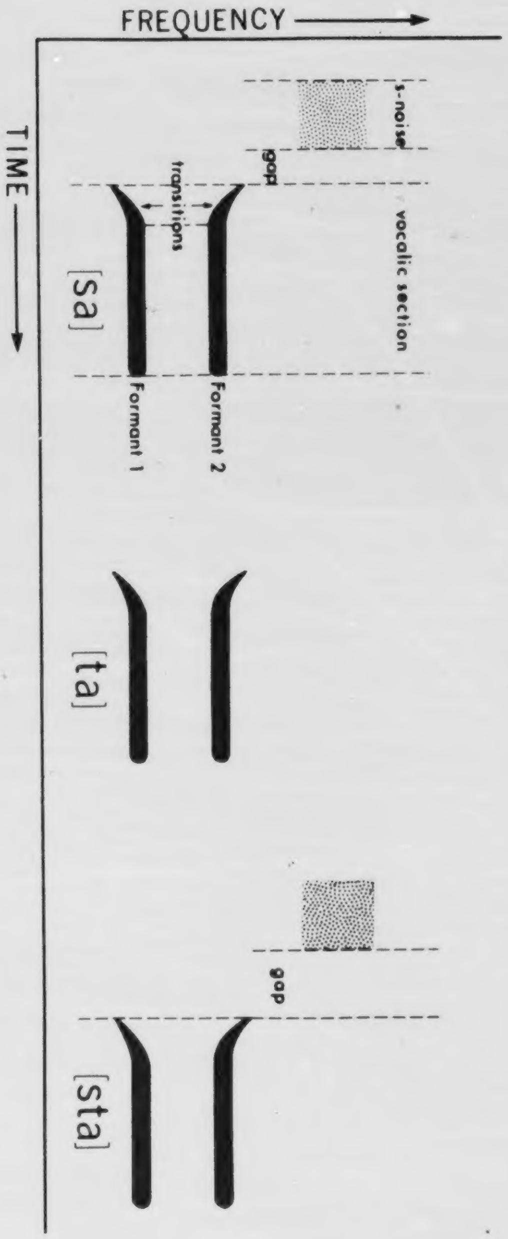

FREQUENCY
TIME
s-noise
gap
transitions
vocalic section
[sa]
Formant 1
Formant 2
[ta]
gap
[sta]

Figure 3: Schematic spectrograms illustrating the importance of silence for the perception of a stop consonant: [sa] becomes [ta] when the noise is removed, or [sta] when a silent interval of approrpiate length is introduced between the noise and the rest of the syllable.

More evidence of the same kind comes from a study of selective adaptation by Ganong (1975). There, the first step was to measure the shift in the (perceived) boundary between [b] and [d] caused by adaptation with the syllable [dε]. Then, a patch of s-noise was placed in front of the [dε] so that it sounded, as in our example, like [sε]. When that syllable ([sε]) was used as the adapter, the effect on the [b-d] boundary was found to be just as great as it had been with [dε]. From that it follows not only that the transition cues were getting through--that is, that they were not being blocked by the noise when they were perceived as [sε] rather than as [dε]--but that they were getting through in full strength.

A third kind of evidence comes from a comparison of how the transition cues are perceived when, in an acoustic context otherwise like that of our example, they are in or out of a proper syllable (Dorman, Raphael, Liberman, and Repp, 1975). The syllable consisted of a patch of s-noise followed by a vocalic portion that was either [pε] or [kε]. With the noise up close, listeners reported hearing [sε], not [spε] or [skε]; [spε] and [skε] were perceived only when there was an appropriate interval of silence between the noise and the rest of the syllable. In the other (nonsyllable, nonspeech) condition, the transition cues were isolated from the rest of the vocalic section, in which circumstance they sounded like 'chirps' of different pitch and could easily be identified on that basis; then they were placed, as in the speech patterns, after the patch of s-noise. In that condition--that is, when heard as 'chirps'--the transition cues were correctly identified even when there was no silent interval separating them from the noise. Thus, they were not significantly masked by the noise, but, just as important from our point of view, their perception was not changed in any qualitative way--that is, there was no apparent interaction among noise, silence, and transitions.

Much the same kind of result has been obtained with stops in syllable-final position (Dorman, et al., 1975). First, it was established that in the disyllables [bεb dε] and [bεg dε] listeners could correctly perceive the syllable-final stops [b] and [g] only if there was a sufficient period of silence (approximately 60 msec) between the syllables. Then, the second-formant transitions that were the only acoustic difference between the [b] and the [g] were isolated from the rest of the pattern of the first syllable, in which circumstance they were heard as two quite different chirps, and presented, as in the first condition, before the syllable [dε]. Listeners correctly identified the chirps most of the time, even when there was no silence at all between them and [dε]; the amount of masking was relatively slight, nothing at all like the total effect that had occurred in the case of the speech sounds, and there appeared, again, to be no interaction-caused change in the phenomenal 'quality'.

So much, then, for the possibility that silence is a necessary condition for perception of stops because it prevents masking of the transitions or because it collaborates in some auditory interaction with them. We turn now to the fact that in the absence of transitions and other stop-consonant cues, silence can be a more nearly sufficient condition for perception of a stop.

Suppose we insert the appropriate amount of silence between the noise of a fricative and a vocalic section so structured that no stop is heard when it is presented by itself. Begin, for example, with the syllable [lit], then put a patch of s-noise in front of it. In that case, the resulting syllable is perceived as [slit] if there is no silence between the noise and the vocalic section, but as [split] if the silence is increased sufficiently (Dorman, Raphael, and Liberman, 1976; Erickson, Fitch, Halwes, and Liberman, 1977). For a simpler example, consider that an appropriate amount of silence inserted between a patch of s-noise and the vowel [i] will produce [ski]; a similar arrangement with [u] will produce [spu] (Summerfield and Bailey, 1977). Notice, too, in these last cases that silence is not only a sufficient cue for stop consonant manner but that the 'place' of the perceived stop (whether [k] or [p]) is different, of which more later.

Silence has also been shown to be a sufficient condition for distinguishing fricative from affricate both in syllable-initial and syllable-final positions. Thus, one can record the word 'say' and the word 'shop' and then convert between 'say shop' and 'say chop' by varying the interval of silence between the two words (Dorman, et al., 1976). Or one can record 'dish' and convert it to 'ditch' by introducing an appropriate amount of silence between the vocalic part of the syllable and the fricative noise at the end.[6]

The foregoing considerations all imply that the perception of silence in our simple example is not only auditory but also phonetic. As a phonetic percept, it conforms to a fact about the speaker's production--namely, that a stop consonant cannot be produced without closing the vocal tract. Of course, such conformity could occur only if there were a phonetic perceiving device specialized to make use of the information about articulation, and if there were, correspondingly, a phonetic mode of perception.

Equivalence in phonetic perception of different acoustic cues produced by the same articulatory gesture. It is a commonplace of speech and speech perception that different acoustic cues may have equivalent effects in phonetic perception. That is of interest because the cues are often so different acoustically that it is hard to conceive how they might be related from an auditory point of view. The relevant facts fall into several classes; we will here offer samples of each.

Perhaps the simplest class comprises those ubiquitous cases in which there are multiple (and distributed) acoustic consequences of the same articulatory gesture. Consider again the example of the preceding section that is owed to Summerfield and Bailey: an appropriate interval of silence between a patch of s-noise and the vowel [i] (or [u]) causes the listener to hear [k] in [ski] (or [p] in [spu]). We now represent that fact schematically in the top half of Figure 4. In the bottom half we represent the companion fact, uncovered in the earlier research on the 'locus' of the stops, that a rising transition at the beginning of the first formant of [i] (or [u]) will also cause a listener to hear the stop [k] in [ki] or ([p] in [pu]) (Delattre, Liberman, and Cooper, 1955). Now we note the perceptual equivalence of 60 or so msec of silence, which is the cue in the top half of the figure, and the rising frequency modulation at the beginning of the first formant, which is the cue in the bottom half, and we ask what that amount of silence and that kind of sound could possibly have in common. Nothing, we should think, when we consider them from an auditory point of view, but in articulation they have an obvious bond. To say [ski] (or [spu]), rather than [si] (or [su]), the speaker must close his vocal tract, which produces the silent interval; and then he must open it, which produces the rise in frequency of the first formant. Thus, the two very different cues are the distributed acoustic results of an essential component of the stop-consonant gesture. Given that they sound alike--either can produce the perception of stop consonant--we should suppose it is because they refer to the same articulation.

---

[6] Raphael and Dorman: personal communication.

For this same example, it remains to take account of the fact that the perceived stops had two different places of production, velar in [ki] (or [ski]) and labial in [pu] (or [spu]). We note, first, that energy at frequency levels corresponding to the second-formant levels of [i] and [u] is appropriate for closure of the vocal tract at the velar and labial places, respectively. That helps us to understand why [i] becomes [ki] (or [ski]) and [u] becomes [pu] (or [spu]) when sufficient cues for the stop manner are added. But notice now a fact that is more relevant to our present purposes, which is that these differences in perception of place of production occur in the same way regardless of how the manner dimension was signaled. Thus, our two very different acoustic cues--silence and sound--are equivalent, not only in their ability to produce the perception of manner, but also in the way they combine with the other information in the signal to produce the perception of phonetic place.

Given our assumption of a link between production and perception, and given that a linguistically significant gesture almost always has multiple acoustic consequences, we should expect to find many other instances of phonetic-perceptual equivalence among cues that are very different in acoustic-auditory terms. Just how many must depend on how finely we dissect the acoustic signal into separate cues, and how often, in experiment, we play the cues off against each other. Relevant studies have already made an impressive record. It reaches back in time to an extension by Lisker (1957b) of an earlier study (Lisker, 1957a) on the voicing distinction in poststress position (as in 'rabid' vs. 'rapid'). Having determined in the earlier work that duration of intersyllabic silence is an important voicing cue, Lisker then found that specifiable amounts of that temporal cue could be traded for specifiable settings of spectral cues (extent of appropriate transitions of the first formant at the end of the first syllable and the beginning of the second). Now, in a recent experiment on the distinction between fricative-vowel and fricative-stop-vowel, Summerfield and Bailey (1977) have established and precisely measured the equivalence of silence on the one hand, and, on the other, such spectral cues as the frequency at which the first formant starts and the extent of the first-formant transition.

There is also evidence of equivalence in phonetic perception among different kinds of temporal cues. Referring again to Lisker's experiment, we note his finding of an equivalence between duration of intersyllable silence and the duration of the first syllable of the word. In a recent experiment,[7]

---

[7] Raphael and Dorman: personal communication.

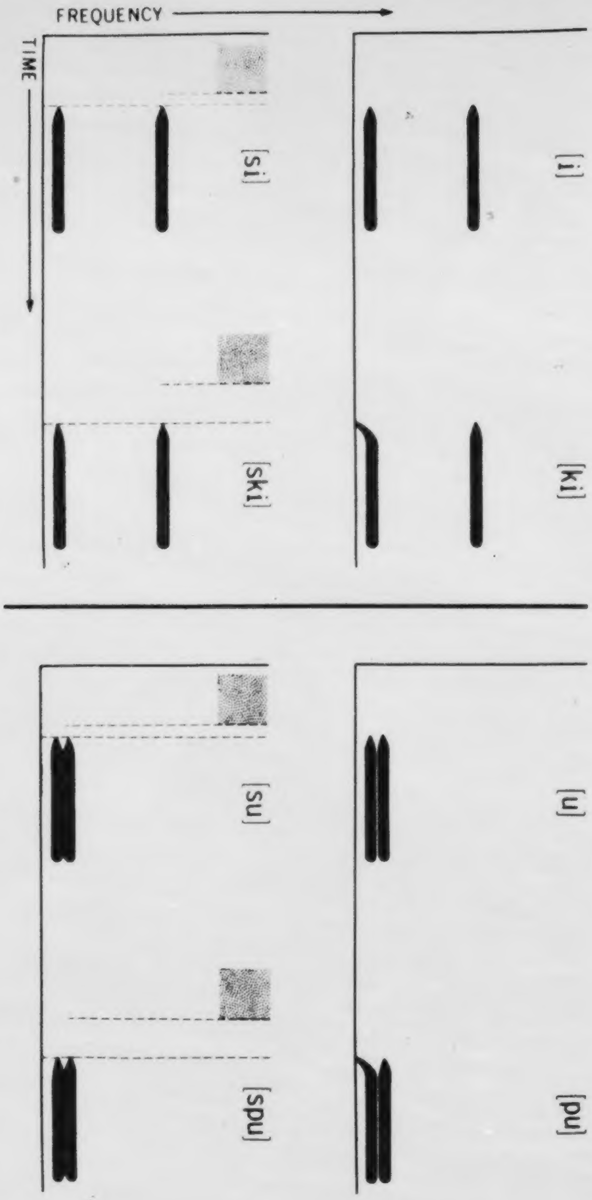

FREQUENCY
TIME
[si]
[ski]
[i]
[ki]
[su]
[spu]
[u]
[pu]

Figure 4: Spectrographic patterns that illustrate how two very different acoustic cues--a transition of the first formant (top half) and an appropriate interval of silence (bottom half)--are phonetically equivalent in the perception of stop consonants.

referred to earlier, on the distinction between [dish] and [ditch] there is an equivalence between the duration of silence separating the vocalic position of the syllable from the noise and the duration of the noise portion of the fricative (or affricate). Also new is the discovery of a similar equivalence between duration of silence and duration of noise in the contrast between fricative-vowel and fricative-stop-vowel.[8] In all these cases time is traded for time; but in the one period of time there is silence, in the other sound.

In the spectral domain, too, equivalences among different cues are not hard to find. For example, an early paper (Cooper et al., 1952) presented preliminary evidence for the separate contributions of several acoustic cues to the perception of the [m-ℓ] distinction, among others. Later, it was shown more clearly that, in the perception of place of production in stops, second- and third-formant transitions made independent contributions (Harris, Hoffman, Liberman, Delattre and Cooper, 1958; Hoffman, 1958; see also Dorman, Studdert-Kennedy and Raphael, in press). In the current literature is a particularly elegant study of the voicing distinction by Summerfield and Haggard (1977) that reports an equivalence between the starting point of the first formant and the variable known as 'voice-onset-time' and shows explicitly how these acoustically disparate cues are related in articulation. A somewhat similar result with two voicing cues--frequency of the fundamental frequency and voice-onset-time--has been found recently by Massaro and Cohen (1976) [cf. Haggard, Ambler, and Callow, (1970)], though an articulatory basis was not made explicit.

Having offered several examples of the equivalences in phonetic perception between different acoustic cues that are the consequences of the same articulation, we should bring this section to a close. But not without first saying that it is hard to know where the list of relevant examples should end. Should we, for example, include the kind of equivalence that is found between spectral cues for syllable-initial consonants and the duration of the syllable,[9] or between silence as a cue (for voicing, or place, or gemination) and the tempo of the surrounding speech (Pickett and Decker, 1960; Port, 1976), or between the setting of the second-formant transition as a cue for the stops and the position of the first formant (Rand, 1971)?[10] It is when we try to answer that question, and thus to define the boundaries of the phenomenon we are here considering, that we see most clearly how unsatisfactory from a theoretical point of view is the notion of acoustic cue. We find it useful, even necessary, when we want to refer to those pieces of sound that an experimenter varied and found to be effective. But if the cues are to be fitted into a conceptual frame--as something other than items in a list--we should regard them as information about the behavior of a speaker's vocal tract.

So far we have considered only those different acoustic cues that are phonetically equivalent because they are the common products of a single articulatory gesture. These are, perhaps, the least complex and most telling of the instances that imply a link between speech perception and speech production. But they are not the only ones. Equally numerous are the cases in which there is phonetic equivalence between acoustic cues that are very different because the phone they signal is produced in different contexts (Liberman, et al., 1967; but see Stevens, 1975). In these cases, too, we suppose that a common articulation is responsible for that which is common in the perception. Of course, such articulations as these can hardly be identical in all particulars, since they are linked to the gestures for the surrounding phones, and these change, of course, with each new context; the commonality can only be seen in terms of shared components, whether end targets or inferred motor commands. (For relevant discussion, see MacNeilage, 1970). But given such articulatory similarity as there may be, gross differences in acoustic signal can and often do arise with changes in context, primarily as a consequence of coarticulation. It is the more important, then, to give some attention to these context-conditioned variations in the cues because, as we said in an earlier section, coarticulation is the essence of the speech code.

---

[8] Bailey, Summerfield and Dorman: personal communication.

[9] Summerfield: personal communication. Miller and Liberman: personal communication.

[10] Also, Bailey: personal communication.

To illustrate how acoustic cues that vary because of phonetic context are nevertheless equivalent in phonetic perception, we choose an example that shows two kinds of contextual effects, one that depends on variations in the identity of the phone following the target phone, and another having to do with variations in the position of the target phone in the syllable. The example is the pair of syllables [did] and [dud], shown schematically as two-formant approximations in Figure 5, and taken from the results of early experiments on the stops (Delattre, et al., 1955). (These patterns are appropriate, and reasonably sufficient, for synthesizing the intended syllables.) Having noticed that the lower (first) formant is the same in the two cases, we fix attention on the higher (second) one. We see there that, as a consequence of coarticulation, a phonetic alteration limited to the middle (vowel) segment of a consonant-vowel-consonant syllable does not change only the middle portion of the sound; rather, it changes the entire second formant. The transition cues for [d] are therefore in very different positions in the spectrum, being relatively high in frequency for [did] and low for [dud]. Moreover, the transition cues for stops in corresponding positions in the syllable are opposite in direction--for [did] they are rising in initial position and falling in final position, but for [dud] they are falling in initial position and rising in final position. Of course, the inference we would draw from these cases is much the same as that we draw from those in which the context was fixed and the disparate acoustic cues were the products of exactly the same gesture: the cues are presumably interpreted by a phonetic device that acts as if it knew how they were produced. But if the device has that ability, then it can conceivably do more than just 'hear through' the context-conditioned variation in the cues so as to arrive at the canonical form of the phonetic segment; it might also be able to take advantage of the fact that such variation produces a special kind of redundancy in the signal and provides important information about such aspects of the phonetic structure as sequential order, juncture, linguistic stress, and tempo. If so, then the acoustic variation that is produced by articulation (and coarticulation) in different contexts would not be an obstacle to perception but a considerable help and, correspondingly, a most important characteristic of the speech code.

FIGURE 5

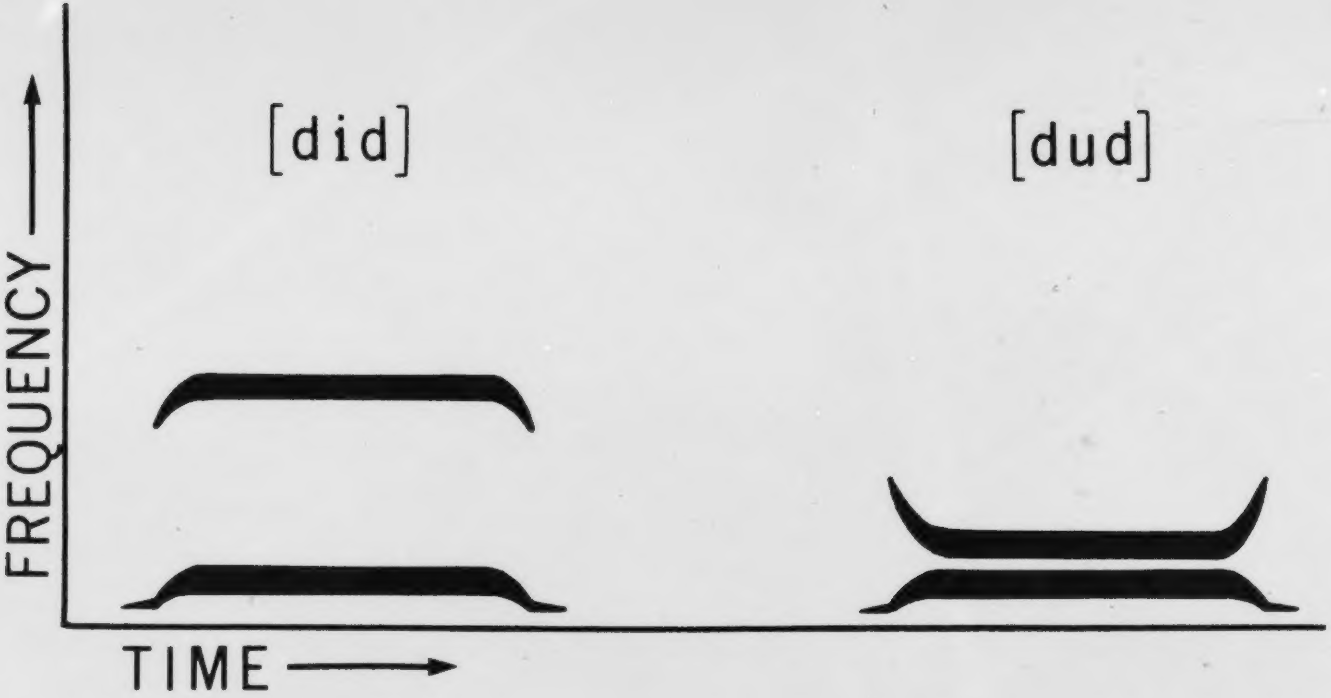


Figure 5: Spectrograms sufficient to produce the syllables [did] and [dud], illustrating the variation in acoustic cues for the stop consonant [d] that can occur as a function of vowel context ([i] vs. [u]) and position in the syllable (initial vs. final).

Non-Equivalence in phonetic perception of an acoustic cue produced by different vocal tracts: ecological constraints of a phonetic sort. Given that phonetic perception is somehow shaped by what a vocal tract does, as we have suggested it might be, we should ask: whose vocal tract? Common sense suggests that it can hardly be that of the listener, nor yet that of the speaker; most plausibly, it must be some abstract conception of vocal tracts in general. We should expect, then, that the phonetic device would behave as if it knew, for example, that two vocal tracts can do what one vocal tract can not. In that case, acoustic cues might have one effect or another, depending on whether they were produced by one speaker or by two. That such ecological considerations are important is indicated by experiments.

One dealt with the perception of the syllable-final stop in the example of [ɛb dɛ] vs. [ɛg dɛ] that we described earlier. There, it will be remembered, listeners could hear the [b] or [g] only if there was a sufficient interval of silence between the syllables, presumably because the phonetic perceiving device 'knew' that the speaker could not have produced both stops without closing his vocal tract for a certain period of time. But two vocal tracts--one saying [ɛb] (or [ɛg]), the other [dɛ]--can produce the disyllable [ɛb dɛ] (or [ɛg dɛ]) with no silence at all between the two syllables. The experiment revealed that listeners behaved accordingly: when a single speaker produced both first and second syllables, a silent interval of some duration was necessary to perception of the syllable-final stops, but when one speaker produced the first syllable and another the second, listeners heard the syllable-final stops even when there was no intersyllabic silence at all. (Dorman, et al., 1975).

The other experiment dealt with the distinction between fricative and affricate (in 'shop' vs. 'chop') that we also described earlier. In that case, inserting a sufficient amount of silence between 'say' and 'shop' caused the listener to hear 'chop'. Our assumption was that this occurred because the silence informed the listener that the speaker had closed his vocal tract, as he must to produce the affricate. But two vocal tracts--one saying 'now say' and the other 'chop'--can produce 'now say chop' with no silence at all between 'say' and 'chop.' Thus, with two speakers, the size of the interval of silence provides no useful phonetic information. The results of the experiment suggested that the listeners' perceptions took account of that fact. Starting with 'now say' and 'shop', and given a silent interval appropriate for 'chop', listeners did indeed hear 'now say chop' if there was only one speaker; but if there were two, then listeners heard 'now say shop' at all intervals of silence (Raphael, Dorman, and Liberman, 1975).

Those results imply that the vocal tract to which the perception is linked is a very abstract one indeed, as we should have expected. But they also provide additional support, and of a rather different kind, for the hypothesis that some such link does indeed exist.

Addition of equivalent acoustic cues: algebraic sums in the phonetic mode. The claim that two very different acoustic cues are equivalent in phonetic perception is largely based on the experimental demonstration of a trading relation between them. Thus, it has been determined that some number of milliseconds of a temporal cue is equal to some particular setting of a spectral cue. An implication is that the two cues together will summate algebraically to enhance or reduce the perceived phonetic contrast, depending on just how they are combined. We believe that to be worth remarking because cues that are algebraically summed would have positive and negative signs only in the phonetic domain, or so it would seem. An example may show why.

Recall the fact, described earlier, that an appropriate period of silence inserted between an s-noise and the syllable [lit] will produce [slit] if the interval is relatively short but [split] if it is sufficiently long. Given that we can, of course, also convert [lit] to [plit] by appropriately changing the spectrum at the beginning of the vocalic syllable--specifically, by altering the formant transitions--it follows that we can use the spectral maneuver to interconvert between [slit] and [split] while holding the temporal cue fixed (Erickson, et al., 1977). Those facts are diagrammed in Figure 6 as Pairs I and II, where we characterize the cues as 'minus p' or 'plus p' to indicate the way they bias the perception. In this case, as in the others we described earlier, we see how a phonetic distinction can be produced by either of two cues, one spectral, the other temporal. In Pairs III and IV, both the spectral and temporal cues differ between the members of the pair, but in different ways. In the one case (Pair III), the combination enhances the perceived difference, while in the other (Pair IV) it permits the minus and plus biases to summate (algebraically) so as to produce two percepts [split] which are the same or very little different (Liberman and Pisoni, in press).

To appreciate the significance of the perceptual addition exemplified in Figure 6, we should think of it as a paradigm for comparative studies with nonhuman animals. Those would be enlightening because phonetic perception, and the algebraic summation that goes with it, exist presumably only in creatures that speak. Others would perceive the stimuli of Figure 6 in an auditory way. Hence, they should find the pairs that differ by two cues (III and IV) to be more discriminable that those (I and II) that differ only by one, and, further, the pairs with two-cue differences should be almost equally discriminable. Note, incidentally, how relatively easy it would be to test that expectation, not only with animals but with human infants: the measure--relative difficulty of discrimination--is surely one of the easiest to make, and the order of difficulty to be expected from nonhuman animals is very different from that already obtained with us human beings.

Non-Equivalence in phonetic perception of the same or similar acoustic cues. Just as the processes of speech production cause different acoustic cues to be correlated in articulation and (hence) equivalent in perception, so also, if in a somewhat more complex way, do they sometimes cause the same cue to be uncorrelated in articulation and (hence) different in perception. An early instance of this was seen in the first 'synthetic' experiment on the stops, where it was found that a burst centered at 1440 Hz was perceived differently in front of different vowels (Liberman, et al., 1952). Subsequently, much the same effect was found with real speech (Schatz, 1954). More recently, the general effect has been confirmed, though with better methods for controlling the stimuli, but now it is seen that the exact nature of the effect varies somewhat depending on just how much of the 'real'

FIGURE 6

	DESCRIPTION OF STIMULI			PERCEPT	CHARACTERIZATION OF CUES			
		GAP	VOCALIC		TEMPORAL	SPECTRAL	TEMPORAL	SPECTRAL
PAIR I	s noise	short	*lit*	*slit*	-p	-p	same	different
		short	*plit*	*split*	-p	+p		
PAIR II	s noise	short	*lit*	*slit*	-p	-p	different	same
		long	*lit*	*split*	+p	-p		
PAIR III	s noise	short	*lit*	*slit*	-p	-p	different	different
		long	*plit*	*split*	+p	+p		
PAIR IV	s noise	short	*plit*	*split*	-p	+p	different	different
		long	*lit*	*split*	+p	-p		

Figure 6: Diagrams that illustrate how spectral and temporal cues separately produce the same phonetic distinction (Pairs I and II) and how, taken together, they either enhance that distinction or reduce it (Pairs III and IV).

burst is used and just where it is placed in time with reference to the vowel.[11]

Another example concerns silence, about which we have already heard so much. Having seen earlier that it is a cue for the perception of phonetic segments, we should note now that it is effective in regard to all three phonetic dimensions: manner, voicing, and place. In connection with manner, we should remember that an appropriate amount of silence, placed between the noise of a fricative and a vocalic piece of sound, will produce the perception of a stop consonant, the perceived 'place' of the stop depending on the nature of the vocalic section. We should also remember that, in similar fashion, silence will produce the affricate manner when introduced appropriately between, for example, the word 'say' and the word 'shop'. In regard to voicing, we saw earlier that variations in the duration of intersyllable silence will convert a poststress voiced stop (as in 'rabid') into voiceless (as in 'rapid'), and vice versa. Now we turn to the dimension of place, and point out, as we had not before, that in a disyllable like 'rabid', reductions in the duration of intersyllable silence will cause the listener to hear 'ratid'--that is, a stop with a different place of production (Port, 1976). This perceptual change correlates--not accidentally, according to our hypothesis--with the fact that a speaker closes his vocal tract for a shorter time when he says 'ratid' than when he says 'rabid'. Given the utterance 'rabid,' and an artificially shortened silence between the syllables, it is as if the listener heard 'ratid' because his phonetic perceiver knows that the speaker could not have said 'rabid' since he did not close his vocal tract long enough. In sum, then, a single acoustic dimension, duration of silence, produces contrasts on each of three phonetic and perceptual dimensions--manner, voicing, and place. That curious situation arises because the very different kinds of articulations--indeed, the different sets of muscles--that underlie the independence of those dimensions in the phonetic and perceptual domains happen to converge on a single acoustic dimension.

Perhaps the reader will have noticed that we did not specify the amounts of silence that are appropriate in the aforementioned cases, and he will quite naturally wonder if they are within the same ranges for the three phonetic and perceptual dimensions. We did not specify because the appropriate durations vary according to how the other relevant cues are set, and much of this remains to be worked out. It is reasonably clear, even now, that the durations of silence for manner and voicing overlap greatly. For place, there probably is some overlap with voicing, depending on just how the other cues are set, but at this moment the relevant data have not been gathered.

Since we have, up to this point, looked only at the segmental aspects of phonetic structure, it may seem inappropriate that we should now broaden our view to glimpse those other aspects that pertain to prosody and syntax. But the temptation to do so is great because there is, at just this juncture, a very natural and interesting connection. The point is that the duration of a

---

[11] Raphael, Dorman and Liberman: personal communication.

syllable conveys information not only about the identity of the phonetic segments it comprises, as we have already seen, but also about the tempo (rate of articulation), degree of linguistic stress, and position in the syntactic frame. We do not know how the separate contributions to duration are sorted out in perception, but, as Klatt (1976) has pointed out, considerations of simple logic suggest that the perceiver can hardly arrive at his decisions in some particular order, one at a time, since each decision would appear to depend on every other one. In any case, it does appear that, in production, these several aspects of the message are encoded into the same aspect of the signal, and then, in perception, properly recovered.

IS PHONETIC PERCEPTION NECESSARY?

So far we have assumed the perceptual reality of phone-size segments. In this final section we propose to justify that assumption.

Preliminary

No one, of course, doubts that linguistic utterances are perceived as sequences of word-like, or morphemic, segments. But the processes by which these segments are extracted from the acoustic signal are far from certain. Do we, in perceiving speech, pass directly from the overall acoustic shape of the constituent morphemes to their syntactic and semantic attributes, or do we, rather, first analyze at least some portion of each utterance (with the possible exception of nonpropositional greetings, interjections, and expletives) into its phonological components, and only then proceed to syntax and meaning? Certainly, the phonological attributes of each morpheme are available to consciousness. But what is the form that gives access to the listener's lexicon? Is lexical storage analog and isomorphic with gross auditory shape, or is it digital and isomorphic with phonological structure?

We should make clear from the outset that the precise form of any possible morphemic sound pattern is not our concern. We have spoken until now of phones, since we take a phonetic representation to be at the first remove from the auditory signal and to be perceptually available, even though not attended to in normal listening. However, for the present discussion, it is a matter of indifference whether the representation is assumed to be a feature matrix, a sequence of phones, or a sequence of more abstract phonologic segments. Our only concern is whether the form is segmented or unsegmented.

Evidence Against Segments Smaller than Syllables

Consider, first, the grounds for believing the perceptual representation to be unsegmented. Foremost is the fact, to which we have repeatedly alluded, that phonetic segments are not discretely arrayed in time, as are letters of the alphabet in space, but are, rather, transmitted simultaneously or with considerable shingling. This fact alone has led some students to abandon the phone as a perceptual unit in favor of the context-sensitive allophone (Wickelgren, 1969; but see Halwes and Jenkins, 1971) or the syllable (Massaro, 1972; Warren, 1976a; 1976b).

A second line of argument draws on reaction-time studies demonstrating that listeners, asked to monitor a word-list or sentence, display successively shorter reaction times as the target item increases in duration from phone to syllable to word (Savin and Bever, 1970) and even to sentence (Bever, 1970), suggesting a perceptual progression from larger unit to smaller rather than the reverse. The solution to this paradox was provided by McNeil and Lindig (1973), who showed that reaction times are, in fact, shortest for the items of which a list is composed or, in other words, for those items to which the experimental situation has drawn the listener's attention. (See also Foss and Swinney, 1973.) Rubin, Turvey and van Gelder (1976), have elaborated these conclusions, arguing that such monitoring experiments do not measure the time taken to process the targets perceptually, but rather the time taken to bring them into consciousness (cf. Studdert-Kennedy, 1974, p. 2366). This does not, of course, preclude the possibility that normal processing entails unconscious access to the lexicon through phonological analysis, since, in all likelihood, these experiments have no bearing on normal perceptual processes at all. However, it does invite the reflection that the several attributes of a morpheme--its phonological components, syllabic structure, syntactic and semantic markers--may all be simultaneously available to the listener, once access to his lexicon has been granted by overall acoustic (or, in reading, visual) shape (Warren, 1976b).

Finally, a broad line of argument springs from the suspicion that the study of speech perception has been tied to the isolated syllable and its components at the expense of attention to the overall acoustic pattern of running speech. This overall pattern, or prosody, certainly conveys important information. Svensson (1974), for example, has shown that the perceived form of hummed speech (that is, speech lacking all the acoustic cues for its phonetic segments) is often syntactically correct. Martin (1972, 1975) has argued that speech rhythm may enable listeners to predict upcoming stresses. And Darwin (1975) has even induced listeners to reveal a preference in some circumstances for good prosody over good syntax and meaning. These and other studies (for example, Cohen and Nooteboom, 1975, passim) do suggest that the role of prosody in speech perception may have been underestimated. In fact, if we combine these studies with recent work on possible invariant acoustic correlates of distinctive features in the speech stream (Stevens, 1975), we may be tempted to propose once again the "novel theory of speech perception," first put forward by Chomsky and Miller (1963, p. 311) and elaborated by Chomsky and Halle (1968, p. 24), by which a few more or less invariant acoustic properties give the listener access to his lexicon and so precipitate a plausible syntactic and semantic analysis of an utterance.

In short, a fair body of evidence suggests that the acoustic structure of spoken utterances may be sufficient to access the listener's lexicon, or at least his syllabary, without an intermediate stage of phonological analysis. However, we do not believe that this view is correct and in the following sections we will try to explain why.

Evidence in Favor of Segments Smaller than the Syllable

Experimental evidence. There is a great weight of evidence for the psychological reality of every level of phonetic analysis, from feature to phone to syllable. We have reviewed much of this evidence elsewhere (Studdert-Kennedy, 1976). Here we do no more than remark that studies of speaking errors (Boomer and Laver, 1968; Fromkin, 1971), perceptual confusions (Miller and Nicely, 1955; Mitchell, 1973), synthetic speech continua (Liberman, et al., 1967), dichotic listening (Shankweiler and Studdert-Kennedy, 1967; Studdert-Kennedy and Shankweiler, 1970) and "verbal transformations" (Goldstein and Lackner, 1973; Warren, 1976a) leave little room for doubt that both phones and features have some form of psychological reality. To this experimental evidence we may add the testimony of linguistic analysis (for example, Gleason, 1955), including studies of language change (for example, Lehman, 1975), not to mention the very existence of alphabetic writing.

The structure of the syllable. As the etymology of its name implies, the syllable is a compound, the vehicle of a natural acoustic contrast between consonant constriction and vowel opening, a contrast frequently claimed as a phonological universal (for example, Postal, 1968). The contrast is clearly reflected in perception, as evidenced by a long series of studies over the past fifteen years. These studies, employing a variety of experimental paradigms--identification and discrimination of synthetic speech sounds, short-term memory, reaction time, dichotic listening, backward masking and others--converge on the conclusion that consonants and vowels perform distinct perceptual functions. Once again, we have reviewed this matter elsewhere (Studdert-Kennedy, 1975b, 1976) and will not do so here. We simply remark that none of the varied evidence for the perceptual contrast between consonants and vowels could exist if the syllable were not analyzed in perception.

We may note, in passing, one further point. In all languages, the syllable is the unit of poetic meter. Except where syllable and morpheme normally coincide (for example, Japanese haiku), metrical rules are specified in terms of syllables and their expected 'length' or degree of stress. Of particular interest in the present context is the fact that length is frequently specified not by lexical form but by neighboring phonetic segments. In both Latin and ancient Greek verse, for example, the length assigned to a word-final CVC syllable varies as a function of the initial phone of the following word. Thus a famous Horace Ode (Book III, Ode XXVI) begins: "Vixi puellis nuper idoneus..." Here the third word scans as a trochee because the following word begins with a vowel, but would have scanned as a spondee had the following word begun with a consonant. This simple rule of ancient verse obviously required that the singer (and presumably the listener who could detect a singer's error) be aware of the phonological structure of the syllable.

The perceptual function of phonological categories. A crucial process in the perception of fluent speech must be short-term storage of early portions of an utterance pending final interpretation. What is the form of this store? Clearly it cannot be simply auditory, since a precategorical auditory store

(Crowder and Morton, 1969; Crowder, 1972) is sensitive to overwriting from immediately following items (Crowder, 1971). Nor, given our sensitivity to phonetic structure (most obviously in listening to poetry) can the store be purely functional or semantic, with all phonetic detail stripped away.

In fact, we wish to argue that, to fulfill this linguistic function, a general perceptual process is invoked, namely, division into 'stages.' Among the likely functions of 'perceptual stages'--whether defined in time or in neural locus--is to isolate one process from another, and to store energy or information for later use. We may see this most clearly at the periphery. Every sensory system integrates energy: if the system were infinitely damped, threshold for activation would never be reached. Accumulation of energy over some finite period permits the mechanical response of the ear, for example, to develop. On the other hand, the period of integration must be finite to prevent physical destruction of the system: mechanical energy becomes bioelectricity. Analogous cycles of integration and transformation presumably recur, as energy or information progresses through the system. Activity in afferent fibers gives rise to more central neural activity and, ultimately (jumping levels of discourse), to a preperceptual 'image' (Massaro, 1972). The 'image', in turn, must have some finite duration, long enough to institute further processing, short enough to prevent 'babble.'

Returning with this metaphor to language, we note that speech is arrayed in time, and that both syntax and meaning demand some minimum quantity of information before linguistic structure can emerge. The perceptual function of phonological categories may then be, on the one hand, to forestall auditory babble, on the other, to store information derived from the signal until such time as it can be granted a linguistic interpretation. In other words, the perceptual function of phonological categories is that of a buffer between acoustic signal and meaningful message.

Recovery of the morpheme. We come, finally, to the phonological function without which, we believe, linguistic communication would not be possible, namely, to provide a code for lexical storage.

Notice first that if lexical items are coded according to overall acoustic structure, the form must be sufficiently stylized, stripped of acoustic detail, for the word to be accessed, despite a wide variety of surface forms. For example, the duration of a single monosyllabic word, spoken by a single speaker at a conversational rate in a random list or in a sentence, may vary by a factor of 2 to 1 (Gaitenby, 1965; Kozhevnikov and Chistovich, 1965; Lackner and Levine, 1975), and yet be fully intelligible in both contexts. Furthermore, the durational variants are not related by a simple scale-factor: most of the variation occurs over the syllable nucleus rather than over its edges (Gaitenby, 1965; Lehiste, 1970; Huggins, 1972), so that an algorithm for generalizing two extreme acoustic variants could hardly succeed without at least some analysis of the overall acoustic shape.

If we add to durational variations, other within-speaker variations in fundamental frequency (which, coupled with duration, is the primary acoustic correlate of variations in linguistic stress) and in formant structure (due to cross-morphemic effects of coarticulation in running speech), not to mention acoustically similar across-speaker variations due to age, sex and dialect, we are confronted with a formidable array of acoustic forms each of which--if unanalyzed acoustic structure is to give access to the lexicon--will have to be reduced to canonical acoustic form.

Now, it is true that the invariance problem is scarcely less serious if the message units to be recovered from the signal are phonological entities such as features or phones than if they are morphemes or words, and even a cursory survey of the literature of speech perception will show that, as in the earlier sections of our chapter, this is a recurrent preoccupation. However, we should note that the 'audile' listener, consigned to lexical search with nothing but overall acoustic shape (and a few syntactic-semantic hints derived from prosody and context) to guide him, is deprived of at least one valuable aid, namely the systematic phonological and phonotactic constraints of his language. He will not be permitted to resolve uncertainty by drawing on his knowledge that a particular portion of the acoustic pattern, or a particular sequence of acoustic segments, cannot occur in his language. Rather, every morphemic sound pattern will be distinct, and access to its semantic and syntactic attributes will be direct. In other words, the vast and subtle array of systematic phonology that linguistic studies have brought into view over the past one hundred and fifty years will be no more than epiphenomenal froth, communicatively vacuous, at least for the listener, if not for the speaker.

Nonetheless, let us set the problem of invariance aside. Let us assume, for the moment, that it has been solved and that we are able to specify for every word or morpheme a unique canonical acoustic form apt for every context and every speaker. We shall then be confronted with the deeper problem of how the listener segments an utterance into its constituent morphemes or words.

The heart of the problem is simply that speakers freely coarticulate across word and morpheme boundaries. A consequence is that dividing the speech stream by use of an acoustic (or auditory) criterion will yield segments that bear a random relation (in size) to the words or morphemes. In that circumstance, the audile listener would have to store, not merely the 20,000 or 30,000 canonical auditory patterns that would represent the words in his vocabulary, but rather a number unimaginably greater than that (Liberman and Pisoni, in press). Even if he had a reliable acoustic criterion for dividing an utterance into syllables (see Mermelstein, 1975), he would not be able to assign the syllables to their appropriate morphemes without analyzing them into their phonetic segments. For example, syllabification of the simple phrase, "He's a repeated offender," will yield eight CV syllables, four of which cross morpheme boundaries and two of which cross word boundaries. In other words, syllable boundaries in fluent speech are frequently random with respect to words or morphemes.

The problem is exacerbated for inflectional languages where changes in a single phoneme (initial, medial or final) often suffice to signal changes in tense, mood, person, number or case. Simple suffix changes, such as English plurals, might pose no problem for the 'audile' listener, despite the lawful

[s], [z] and [ɨz] alternations, and the absence of an acoustically marked morpheme boundary, for we need only suppose that the perceptual 'morpheme detector' is automatically sprung as soon as a recognizable acoustic unit enters the system. We might even suppose that tense contrasts signalled by a change in medial vowel (as in 'win'-'won') are learned as special cases. But we will find it a good deal more difficult to explain, for example, the formation of the Greek perfect tense by duplication of the initial consonant of the present, a fact presumably not lost on the listener. Indeed, as we multiply examples (and ad hoc solutions for the imaginary 'audile' listener), we cannot but wonder why the various forms of a lexical item bear any relation to one another at all. Are we to suppose that these variations are lawful for the speaker, but merely adventitious for the listener?

Surely not. For, quite apart from the general lack of parsimony in positing totally independent input and output lexicons, we would be reduced to the absurdity of supposing that a listener consults a lexicon of auditory segments that bear no more than a random relation to the articulatory segments he deploys as a speaker. We are forced to conclude that only by extracting the phonetic segments--or, more properly, their underlying phonological forms--can the listener discover most of what is said to him.

REFERENCES

Ainsworth, W. A. (1976) Automatic speech recognition. In *Mechanisms of Speech Recognition*. (Oxford, Eng.: Pergamon Press), 104-119.

Ades, A. E. (in press) Source assignment and feature extraction in speech. *J. Exp. Psychol.: Human Percept. Perform.*

Ades, A. E. (1976) Adapting the property detectors for speech perception. In *New Approaches to Language Mechanisms*, ed. by E. C. T. Walker and R. J. Wales. (Amsterdam: North Holland).

Bailey, P. J. (1973) Perceptual adaptation for acoustical features in speech. *Speech Perception*, Progress Report. (Department of Psychology, The Queen's University of Belfast), *2.2*, 29-34.

Bailey, P. J., M. F. Dorman and A. Q. Summerfield. (1977) Identification of sine-wave analogues of CV syllables in speech and non-speech modes. *J. Acoust. Soc. Am. 61*, S(A).

Bever, T. G. (1970) The influence of speech performance on linguistic structure. In *Advances in Psycholinguistics*, ed. by G. B. Flores D'Arcais and W. J. M. Levett. (Amsterdam: North Holland).

Boomer, D. S. and J. D. M. Laver. (1968) Slips of the tongue. *Brit. J. Dis. Communic. 3*, 1-12.

Chomsky, N. and M. Halle. (1968) *The Sound Pattern of English*. (New York: Harper & Row).

Chomsky, N. and G. A. Miller. (1963) Introduction to the formal analysis of natural languages. In *Handbook of Mathematical Psychology*, vol. I, ed. by R. D. Luce, R. R. Bush and E. E. Galauker. (New York: Wiley), 269-321.

Cohen, A. and S. G. Nooteboom, eds. (1975) *Structure and Process in Speech Perception*. (New York: Springer-Verlag).

Coker, C., N. Umeda and C. P. Browman. (1973) Automatic synthesis from ordinary English text. *IEEE Trans. Audio Electroacoust. 21*, 293-298.

Cole, R. A. and B. Scott. (1974) Toward a theory of speech perception. *Psych. Rev. 81*, 348-374.

Cooper, F. S. (1950) Spectrum analysis. *J. Acoust. Soc. Amer. 22*, 761-762.

Cooper, F. S. (1953) Some instrumental aids to research on speech. In *Proceedings of the Fourth Annual Round Table Meeting on Linguistics and Language Teaching*. (Washington: Georgetown Univ.), 46-53.

Cooper, F. S. (1962) Speech synthesizers. In *Proceedings of the Fourth International Congress of Phonetic Sciences*, ed. by A. Sovijatvi and D. Aalto. (The Hague: Mouton), pp. 3-13.

Cooper, F. S. (1972) How is language conveyed by speech? In *Language by Ear and by Eye*, ed. by J. F. Kavanagh and I. G. Mattingly. (Cambridge, Mass.: MIT Press), 25-45.

Cooper, F. S., P. C. Delattre, A. M. Liberman, J. M. Borst and L. J. Gerstman. (1952) Some experiments on the perception of synthetic speech sounds. *J. Acoust. Soc. Am. 24*, 597-606.

Cooper, F. S., A. M. Liberman, and J. M. Borst. (1951) The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Sciences 37*, 318-325.

Cooper, W. E. (1975) Selective adaptation to speech. In *Cognitive Theory*, vol. 1, ed. by F. Restle, R. M. Shiffrin, J. N. Castellan, H. Lindman and D. B. Pisoni. (Hillsdale, N.J.: Erlbaum Associates), 23-54.

Crowder, R. G. (1971) Waiting for the stimulus suffix: Decay, delay, rhythm and readvent in immediate memory. *Quart. J. Exp. Psychol. 23*, 324-340.

Crowder, R. G. (1972) Visual and auditory memory. In *Language by Ear and by Eye*, ed. by J. F. Kavanagh and I. G. Mattingly. (Cambridge: MIT Press).

Crowder, R. G. and J. Morton. (1969) Precategorical acoustic storage (PAS). *Percept. Psychophys. 5*, 365-373.

Darwin, C. J. (1971) Ear differences in the recall of fricatives and vowels. *Quart. J. Exp. Psychol. 23*, 46-62.

Darwin, C. J. (1975) On the dynamic use of prosody in speech perception. In *Structure and Process in Speech Perception*, ed. by A. Cohen and J. G. Nooteboom. (New York: Springer-Verlag), 178-194.

Darwin, C. J. (1976) The Perception of Speech. In *Handbook of Perception*, vol. 7, ed. by E. Carterette and M. Friedman. (New York: Academic Press), pp. 175-226.

Day, R. S. (1970) Temporal order perception of reversible phoneme cluster. *J. Acoust. Soc. Am. 48*, 95(A).

Delattre, P. C., A. M. Liberman and F. S. Cooper. (1955) Acoustic loci and transitional cues for consonants. *J. Acoust. Soc. Am. 27*, 769-773.

Denes, P. B. (1965) Effect of duration on the perception of voicing. *J. Acoust. Soc. Am. 27*, 761-764.

Dorman, M., J. E. Cutting and L. J. Raphael. (1975) Perception of temporal order in vowel sequences with and without formant transitions. *J. Exp. Psychol.: Human Percept. Perfor. 104*, 121-129.

Dorman, M. F., L. J. Raphael and A. M. Liberman. (1976) Further observations on the role of silence in the perception of stop consonants. *Haskins Laboratories Status Report on Speech Research SR-48*, 199-207.

Dorman, M. F., L. J. Raphael, A. M. Liberman and B. Repp. (1975) Maskinglike phenomena in speech perception. *J. Acoust. Soc. Am. 57*, *Suppl. 1*, S48(A). [Full text in *Haskins Laboratories Status Report on Speech Research SR-42/43*, 265-276.]

Dorman, M., M. Studdert-Kennedy and L. J. Raphael. (in press) Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Percept. Psychophys.*

Eimas, P. D. and J. D. Corbit. (1973) Selective adaptation of linguistic feature detectors. *Cog. Psychol.* *13*, 247-252.

Erickson, D. M., H. L. Fitch, T. G. Halwes and A. M. Liberman. (1977) Trading relation in perception between silence and spectrum. *J. Acoust. Soc. Am.* *61*, S46(A).

Fant, C. G. M. (1962) Descriptive analysis of the acoustic aspects of speech. *Logos* *5*, 3-17.

Fant, C. G. M. (1968) Analysis and synthesis of speech processes. In *Manual of Phonetics*, ed. by B. Malmberg. (Amsterdam: North-Holland), pp. 173-277.

Foss, D. J. and D. A. Swinney. (1973) On the psychological reality of the phoneme: perception, identification and consciousness. *J. Verbal Learn. Verbal Behav.* *12*, 246-257.

Fromkin, V. A. (1971) The non-anomalous nature of anomalous utterances. *Language* *47*, 27-52.

Fujimura, O. (in press) A look into the effects of context: some articulatory and perceptual findings. In *Proceedings of VIIIth International Congress of Phonetic Sciences*, (Leeds, 1975).

Fujimura, O. (1975) Syllable as a unit of speech recognition. *IEEE Trans. Acoust. Sp. Sig. Proc.* *ASSP-23*, 82-87.

Gaitenby, J. H. (1965) The elastic word. *Haskins Laboratories Status Report on Speech Research SR/2*, 3.1.

Ganong, W. F. (1975) An experiment on "phonetic adaptation." *Progress Report* (Research Laboratory of Electronics, MIT) *116*, 206-210.

Gleason, H. A., Jr. (1955) *Workbook in Descriptive Linguistics*. (New York: Holt, Rinehart & Winston).

Goldstein, L. M. and J. R. Lackner. (1973) Alterations of the phonetic coding of speech sounds during repetition. *Cognition* *2*, 279-297.

Haggard, M. P., S. Ambler, and M. Callow. (1970) Pitch as a voicing cue. *J. Acoust. Soc. Am.* *47*, 613-617.

Halwes, T. and J. J. Jenkins. (1971) Problem of serial order in behavior is not resolved by context-sensitive associative memory models. *Psych. Rev.* *78*, 122-129.

Harris, C. M. (1953) A study of the building blocks of speech. *J. Acoust. Soc. Am.* *25*, 962-969.

Harris, K. S. (1958) Cues for the discrimination of American English fricatives in spoken syllables. *Lang. Speech* *1*, 1-7.

Harris, K. S., H. S. Hoffman, A. M. Liberman, P. C. Delattre, and F. S. Cooper. (1958) Effect of third-formant transitions on the perception of the voiced stop consonants. *J. Acoust. Soc. Am.* *30*, 122-126.

Hoffman, H. S. (1958) Study of some cues in the perception of the voiced stop consonants. *J. Acoust. Soc. Am.* *30*, 1035-1041.

Huggins, A. W. F. (1972) On the perception of temporal phenomena in speech. *J. Acoust. Soc. Am.* *51*, 1279-1290.

Hyde, S. R. (1972) Automatic speech recognition: A critical survey and discussion of the literature. In *Human Communication: A Unified View*, ed. by E. E. David, Jr. and P. B. Denes. (N.Y.: McGraw-Hill), pp. 399-438.

Ingemann, F. (1957) Speech synthesis by rule. *J. Acoust. Soc. Am. 29*, 1255(A).

Kelly, J. L. and L. J. Gerstman. (1961) An artificial talker driven from a phonetic input. *J. Acoust. Soc. Am. 33*, 835(A).

Kelly, J. L. and C. Lochbaum. (1962) Speech synthesis. Proceedings of the Speech Communications Seminar, paper F7. (Stockholm: Speech Transmission Laboratory, Royal Institute of Technology).

Klatt, D. H. (1976) Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. Acoust. Soc. Am. 59*, 1208-1221.

Klima, E. S. (1975) Sound and its Absence in the Linguistic Symbol. In *The Role of Speech in Language*, ed. by J. F. Kavanagh and J. E. Cutting. (Cambridge: MIT Press), pp. 247-270.

Kozhevnikov, V. A. and L. A. Chistovich. (1965) *Rech' Artikuliatsia i vospriatie*, Moscow-Leningrad. Trans. as *Speech: Articulation and Perception*. (Washington: Clearinghouse for Federal Scientific and Technical Information), JPRS, 30.

Lackner, J. R. and B. K. Levine. (1975) Speech production: Evidence for syntactically and phonologically determined units. *Percept. Psychophys. 17*, 107-113.

Ladefoged, P. (1971) *Preliminaries to Linguistic Phonetics*. (Chicago: University of Chicago Press).

Lehiste, I. (1970) Suprasegmentals. (Cambridge: MIT Press).

Lehman, W. (1975) *Historical Linguistics*, 2nd ed. (New York: Holt, Rinehart & Winston).

Liberman, A. M. (in press) How abstract must a motor theory of speech perception be? In *Proceedings of the Eighth International Congress of Phonetic Sciences, Leeds, England, 17-23 August, 1975*. [Also in *Haskins Laboratories Status Report on Speech Research SR-44*, 1-15, (1975).]

Liberman, A. M. (1976) Discussion paper. In *Origins and Evolution of Language and Speech*, ed. by S. R. Harnad, H. D. Steklis, and J. Lancaster (New York: New York Academy of Sciences), 718-724. [*Annals of the New York Academy of Sciences 280*, 718-724 (1976).]

Liberman, A. M. (1974) The specialization of the language hemisphere. In *The Neurosciences: Third Study Program*, ed. by F. O. Schmitt and F. G. Worden. (Cambridge, Mass.: MIT Press), pp. 43-56.

Liberman, A. M. (1970) The grammars of speech and language. *Cog. Psychol. 1*, 301-323.

Liberman, A. M. and D. B. Pisoni. (in press) Evidence for a special speech-perceiving subsystem in the human. In *The Recognition of Complex Acoustic Signals*, ed. by T. H. Bullock. (Berlin: Dahlem Konferenzen).

Liberman, A. M., F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. (1967) The perception of the speech code. *Psychol. Rev. 74*, 431-461.

Liberman, A. M., P. C. Delattre, and F. S. Cooper. (1952) The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Am. J. Psychol. 55*, 497-516.

Liberman, A. M., F. Ingemann, L. Lisker, P. Delattre, and F. S. Cooper. (1959) Minimal rules for synthesizing speech. *J. Acoust. Soc. Am. 31*, 1490-1499.

Liberman, A. M., I. G. Mattingly, and M. T. Turvey. (1972) Language codes and memory codes. In *Coding Processes in Human Memory*, ed. by A. W. Melton and E. Martin. (Washington, D.C.: V. H. Winston), pp.

307-334.

Lisker, L. (1957a) Some cues to the voiced-voiceless distinction among the intervocalic stops in English. *Haskins Laboratories 24th Quarterly Progress Report*, Appendix 5.

Lisker, L. (1957b) Closure duration and the intervocalic voiced-voiceless distinction in English. *Language 33*, 42-49.

MacNeilage, P. F. (1970) Motor control of serial ordering of speech. *Psych. Rev. 77*, 182-196.

Martin, J. G. (1975) Rhythmic expectancy in continuous speech perception. In *Stimulus and Process in Speech Perception*. (New York: Springer-Verlag), pp. 161-177.

Martin, J. G. (1972) Rhythmic hierarchical versus serial structure in speech and other behavior. *Psychol. Rev. 79*, 487-509.

Massaro, D. W. (1972) Preperceptual images, processing time and perceptual units in auditory perception. *Psych. Rev. 79*, 124-145.

Massaro, D. W. and M. M. Cohen. (1976) The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. *J. Acoust. Soc. Am. 60*, 704-717.

Mattingly, I. G. (1976) Syllable synthesis. *J. Acoust. Soc. Am. 60*, S75(A). [Also in *Haskins Laboratories Status Report on Speech Research SR-49* (1977).]

Mattingly, I. G. (1974) Speech synthesis for phonetic and phonological models. In *Current Trends in Linguistics*, Vol. 12: *Linguistics and Adjacent Arts and Sciences*, ed. by T. A. Sebeok, et al. (The Hague: Mouton), 2451-2488.

Mattingly, I. G. (1972) Speech cues and sign stimuli. *American Scientist 60*, 327-337.

Mattingly, I. G. and A. M. Liberman. (1969) The speech code and the physiology of language. In *Information Processing in the Nervous System*, ed. by K. N. Leibovic. (New York: Springer-Verlag), pp. 97-117.

Mattingly, I. G., A. M. Liberman, G. K. Syrdal, and T. Halwes. (1971) Discrimination in speech and nonspeech modes. *Cog. Psychol. 2*, 131-157.

McNeill, D. and L. Lindig. (1973) The perceptual reality of phonemes, syllables, words and sentences. *J. Verbal Learn. Verbal Behav. 12*, 419-430.

Mermelstein, P. (1975) Automatic segmentation of speech into syllabic units. *J. Acoust. Soc. Am. 58*, 880-883.

Miller, G. A. and P. Nicely. (1955) An analysis of some perceptual confusions among some English consonants. *J. Acoust. Soc. Am. 27*, 338-352.

Miller, J. D. (in press) Perception of speech sounds by animals: Evidence for speech processing by mammalian auditory mechanisms. In *Recognition of Complex Acoustic Signals*, ed. by T. H. Bullock (Berlin: Dahlem Konferenzen 1977).

Mitchell, P. D. (1973) A test of differentiation of phonemic feature contrasts. Unpublished Ph.D. Dissertation, City University of New York.

Ochiai, K. and O. Fujimura. (1971) Vowel identification and phonetic contexts. (Report of University of Electro-Communications, Chofu, Tokyo), *22-2*, 103-111.

Ogden, C. K. (1967) *Opposition*. (Bloomington: Indiana University Press).

Peterson, G. E. and H. L. Barney. (1952) Control methods used in a study of vowels. *J. Acoust. Soc. Am. 24*, 175-184.

Peterson, G. E., W. S-Y. Wang, and E. Sivertsen. (1958) Segmentation techniques of speech synthesis. *J. Acoust. Soc. Am.*, *30*, 739-742.

Pickett, J. M. and L. R. Decker. (1960) Time factors in perception of a double consonant. *Lang. Speech 3*, 11-17.

Pisoni, D. B. (in press) Speech Perception. In *Handbook of Learning and Cognitive Processes*, ed. by W. Estes, vol. 6. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Port, R. F. (1976) The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words. Unpublished Ph.D. dissertation, University of Connecticut.

Postal, M. (1968) *Aspects of Phonological Theory*. (New York: Harper & Row).

Rand, T. C. (1971) Vocal tract size normalization in the perception of stop consonants. *Haskins Laboratories Status Report on Speech Research SR-25/26*, 141-146.

Rand, T. C. (1974) Dichotic release from masking for speech. *J. Acoust. Soc. Am. 55*, 678-680.

Raphael, L. J. (1972) Preceding vowel duration as a cue to the perception of voicing characteristic of word-final consonants in American English. *J. Acoust. Soc. Am. 51*, 1296-1303.

Raphael, L. J., M. F. Dorman, and A. M. Liberman. (1975) Perception of vowel and syllable duration in VC and CVC syllables. *J. Acoust. Soc. Am. 58*, S57(A).

Rubin, P., Turvey, M. T., and P. Van Gelder. (1976) Initial phonemes are detected faster in spoken words than in spoken nonwords. *Percept. Psychophys. 19*, 394-398.

Savin, H. B. and T. G. Bever. (1970) The non-perceptual reality of the phoneme. *J. Verbal Learn. Verbal Behav. 9*, 295-302.

Schatz, C. D. (1954) The role of context in the perception of stops. *Language 30*, 47-56.

Shankweiler, D. P. and M. Studdert-Kennedy. (1967) Identification of consonants and vowels presented to left and right ears. *Quart. J. Exp. Psychol. 19*, 59-63.

Shattuck, S. R. and D. H. Klatt. (1976) The perceptual similarity of mirror-image acoustic patterns in speech. *Percept. Psychophys. 20*, 470-474.

Stevens, K. N. (1975) The potential role of property detectors in the perception of consonants. In *Auditory Analysis and Perception of Speech*, ed. by C. G. M. Fant and M. A. A. Tatham. (New York: Academic Press), 303-330.

Stevens, K. N. and A. S. House. (1972) The perception of speech. In *Foundations of Modern Auditory Theory*, ed. by J. Tobias. (New York: Academic Press), pp. 3-62.

Strange, W., R. Verbrugge, D. P. Shankweiler, and T. R. Edman. (1976) Consonant environment specifies vowel identity. *J. Acoust. Soc. Am. 60*, 198-212.

Studdert-Kennedy, M. (1974) The perception of speech. In *Current Trends in Linguistics*, ed. by T. A. Sebeok, vol. 12 (4). (The Hague: Mouton), pp. 2349-2385.

Studdert-Kennedy, M. (1975a) From acoustic signal to phonetic message. *J. Communic. Dis. 8*, 181-188.

Studdert-Kennedy, M. (1975b) From continuous signal to discrete message: Syllable to phoneme. In *The Role of Speech in Language*, ed. by

J. F. Kavanagh and J. E. Cutting. (Cambridge: MIT Press), pp. 113-125.

Studdert-Kennedy, M. (1976) Speech perception. In *Contemporary Issues in Experimental Phonetics*, ed. by N. J. Lass. (New York: Academic Press), pp. 243-293.

Studdert-Kennedy, M. (in press) Universals in phonetic structure and their role in linguistic communication. In *The Recognition of Complex Acoustic Signals*, ed. by T. H. Bullock. (Berlin: Dahlem Konferenzen).

Studdert-Kennedy, M. and D. P. Shankweiler. (1970) Hemispheric specialization for speech perception. *J. Acoust. Soc. Am. 48*, 579-594.

Summerfield, A. Q. and P. J. Bailey. (1977) On the dissociation of spectral and temporal cues for stop consonant manner. *J. Acoust. Soc. Am. 61*, S(A).

Summerfield, A. Q. and M. P. Haggard. (1977) On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Haskins Laboratories Status Report on Speech Research SR-49*, 1-36.

Svennson, S-G. (1974) Prosody and grammar in speech perception. *MILUS 2*. (Institute of Linguistics, University of Stockholm).

Warren, R. M., C. J. Obusek, R. M. Farmer, and R. P. Warren. (1969) Auditory sequence: Confusion of patterns other than speech and music. *Science 164*, 586-587.

Warren, R. M. (1976a) Auditory sequence and classification. In *Contemporary Issues in Experimental Phonetics*, ed. by N. J. Lass. (New York: Academic Press), pp. 389-417.

Warren, R. M. (1976b) Auditory perception and speech evolution. In *Origins and Evolution of Language and Speech*, ed. by S. R. Harnad, H. D. Steklis, and J. Lancaster. (New York: New York Academy of Sciences), 708-717. [*Annals of the New York Academy of Sciences 280*, 708-717.]

Whitfield, I. C. (1965) 'Edges' in auditory information processing. In *Proceedings of XXIIIrd International Congress of Physiological Sciences*, (Tokyo), September, pp. 245-247.

Whitfield, I. C. and E. F. Evans. (1965) Responses of auditory cortical neurons to stimuli of changing frequency. *J. Neurophysiol. 28*, 655-672.

Wickelgren, W. A. (1969) Context-sensitive coding, associative memory and serial order in (speech) behavior. *Psychol. Rev. 76*, 1-15.

Cardiac Indices of Infant Speech Perception: Orienting and Burst Discrimination*

Cynthia L. Miller†, Philip A. Morse† and Michael F. Dorman††

ABSTRACT

The present study investigated burst cue discrimination in 3- to 4-month-old infants with the natural speech stimuli [bu] and [gu]. The experimental stimuli consisted of either a [bu] or a [gu] burst attached to the formants of the [bu], such that the sole difference between the two stimuli was the initial burst cue. Infants were tested using a cardiac orienting response (OR) paradigm that consisted of 20 tokens of one stimulus (for example, [bu]) followed by 20 tokens of the second syllable (20/20 paradigm). An OR to the stimulus change revealed that young infants can discriminate burst cue differences in speech stimuli. Discussion of the results focused on asymmetries observed in the data and the relationship of these findings to our previous failure to demonstrate burst discrimination using the habituation/dishabituation cardiac measure generally employed with older infants.

---

*Portions of this paper were presented at the biennial meetings of the Society for Research in Child Development, Denver, Colorado, April, 1975. This research is part of a thesis prepared by the senior author in partial fulfillment of the M.A. degree in the Department of Psychology, University of Wisconsin-Madison. This paper is in press in the Quarterly Journal of Experimental Psychology.

†University of Wisconsin, Madison, Wisconsin.

††Also Arizona State University, Tempe, Arizona.

Acknowledgment: This research was supported by NICHD Grants HD-01994, 1-P01HD 80240-01, HD-03352, and PHS Training Grant 5 T01 HD-00117. The authors are indebted to J. Brown, J. Hirsch, P. Hirsch, C. Gillman, F. Graham, P. Buckley and G. Nielson for their help in data analysis, and to N. Cowan, P. Skemp, H. Hollenbeck, D. Horowitz, P. Roth, K. Wyatt and T. Clees for their invaluable assistance in testing subjects and behavior recording. The authors would also like to thank Drs. A. Liberman and F. Cooper for making the facilities of Haskins Laboratories available for the generation of the stimulus tapes.

[HASKINS LABORATORIES: Status Report on Speech Research SR-50 (1977)]

INTRODUCTION

The perception of speech involves the integration of overlapping acoustic cues in the form of a complex code, the primary unit of which has been defined as the syllable (Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967; Liberman, 1970). In syllables consisting of a stop consonant plus a vowel, a brief, initial burst of energy, and (in some stops) aspiration, typically precede the formant transitions and steady-state vowel components of the syllable. Although the importance of all of these components in adult speech perception has been investigated, perhaps the least studied and most controversial of these components has been the initial burst cue.

The burst cue consists of (i), a brief explosion (less than 20 msec) produced by the release of occlusion, and (ii) a very brief (0-10 msec) period of frication. The duration and frequency characteristics of the burst vary as a function of place of articulation, vowel context and voicing. The perceptual significance of the burst also varies as a function of these factors. For example, bursts contribute little to the identification of /b/ except, possibly, in back vowel environments; bursts contribute significantly to the identification of /d/ in front vowel environments, but relatively less in center and back vowel environments; and bursts are generally important for identification of /g/ (for details see Fischer-Jørgensen, 1972; Dorman, Studdert-Kennedy and Raphael, in press). Bursts, then, are important cues for stop consonant recognition.

In a recent study by Morse, Leavitt, Miller and Romero (in press), adult burst discrimination was investigated using a nonverbal, cardiac measure. In this study a heart-rate (HR) orienting response (OR) habituation-dishabituation procedure was used in assessing the discrimination of a natural [bu] ([bu] burst + [bu] formants) and a transposed [gu], consisting of a [gu] burst attached to the same [bu] formants. In the type of paradigm employed by Morse et al., the subject is presented with repeated trials of a familiarization stimulus followed by 1 or 2 trials of a change stimulus. To allow for recovery of a cardiac response to trial offset, the intertrial intervals (ITI) in this paradigm typically vary between 25 and 60 secs. In the Morse et al. study, subjects received 8 trials [each consisting of 8 stimuli with an intersimulus interval (ISI) of 1 sec] of the familiarization stimulus (either [bu] or [gu]) followed by 2 trials of the change stimulus. Intertrial intervals (offset to onset) varied randomly between 25 and 35 sec. In this study, habituation of the cardiac component of the orienting response (HR deceleration) to trial onset was observed across the familiarization trials. Dishabituation (recovery of the orienting response) was found to occur in response to the onset of the change trials, thereby indicating discrimination of those burst cues. In addition, this cardiac evidence of discrimination was accompanied by verbal reports of discrimination.

If bursts do play such an important role in adult speech perception, then it is of interest to examine the developmental course of their importance beginning in early infancy. Research on infant speech perception has shown that by four months of age, infants can already discriminate formant transition and steady-state vowel information in a manner similar to the adult's perception of these cues (for example, Eimas, Siqueland, Jusczyk and Vigorito,

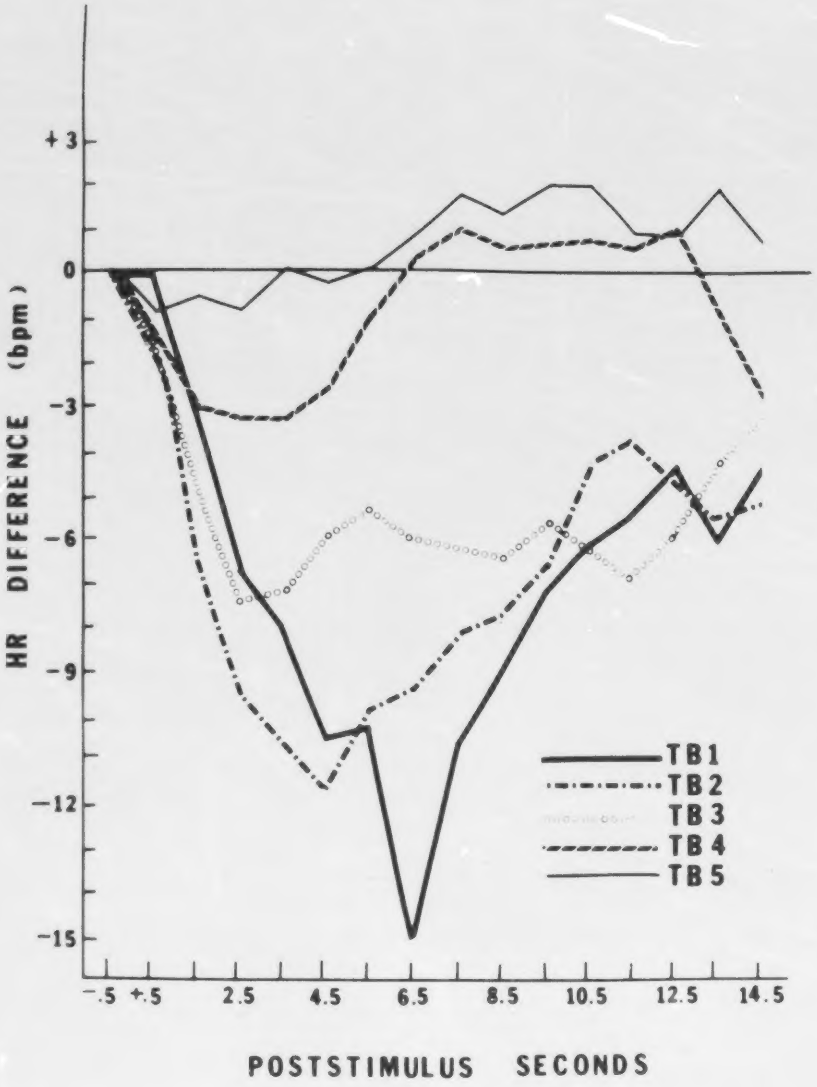


Figure 1: HR difference score data from the 8/2 procedure. Ten trials are grouped into 5 trial blocks of 2 trials each (Miller, et al., 1975a).

1971; Eimas, 1974, 1975a; Miller, 1974; Miller and Morse, 1976; Swoboda, Morse and Leavitt, 1976; Till, 1976). In contrast, no comparable data are yet available on the infant's ability to discriminate differences in the burst component of the syllable. Since variations of the cardiac habituation-dishabituation paradigm described above have been successively employed in several studies of infant auditory discrimination (Moffitt, 1971; Berg, 1972; Lasky, Syrdal-Lasky and Klein, 1975), an earlier study (Miller, Morse and Dorman, 1975) attempted to investigate infant burst cue discrimination using the same heart rate procedure and stimuli employed in the Morse et al. (in press) adult study. The 3- to 4-month old infants in this study (hereafter referred to as Miller et al., 1975a) were presented with 8 trials of either [bu] or [gu] and 2 trials of the change stimulus. The cardiac data from this study are depicted in Figure 1. In this figure, the 10 trials of the 8/2 procedure are grouped into 5 trial blocks (TB) of 2 trials each. Analyses of variance for trends over seconds and trends over trials performed on these data confirmed the observation of a reliable orienting response (OR) to trial onset on the first few trials (TB 1-3) that subsequently habituated over trials. However, dishabituation on the change trials (TB 5) was not observed, thus suggesting that these infants were not capable of discriminating this burst contrast (for additional details of the methodology and results of this study, cf. Miller et al., 1975). However, recent developmental studies of infant auditory discrimination suggest that this conclusion may be premature. Although infants older than 4 months of age have been found to readily exhibit auditory discrimination with variations of the 8/2 cardiac OR paradigm (Moffitt, 1971; Berg, 1974; Lasky, et al., 1975), infants between 6 weeks and 12 weeks have failed to demonstrate auditory discrimination when these paradigms were employed (Berg, 1974; Brown, Leavitt, and Graham, 1975; in press; Leavitt, Brown, Morse and Graham, in press). For example, Brown et al. (1975) recently employed a 6/2 cardiac paradigm (6 familiarization trials, 2 novel trials) in assessing the discrimination of an auditory change in 12-week old infants. As in the Miller et al. (1975a) study, Brown et al. (1975) observed significant orienting and habituation, but no dishabituation to a stimulus change.

In contrast to the failure of infants less than 4 months of age to evidence auditory discrimination using an habituation-dishabituation cardiac paradigm, several investigators have reported auditory/speech discrimination in infants as young as 4 weeks of age using an operant high-amplitude sucking paradigm (Eimas, Siqueland, Jusczyk and Vigorito, 1971; Trehub and Rabinovitch, 1972). Consequently, the absence of discrimination in the present experiment (and in the Brown et al., 1975, study) may more likely reflect a developmental limitation of the habituation-dishabituation paradigm, rather than an inability of young infants to discriminate burst cues. A recent study by Leavitt et al. (in press) further supports this suspicion. Leavitt et al. failed to obtain auditory discrimination in 6-week olds using a 6/2 paradigm, but when a cardiac paradigm was employed in which the intertrial intervals of the habituation/dishabituation procedure were eliminated (a no-ITI paradigm), 6-week olds did exhibit auditory discrimination.

In sum, this developmental pattern of heart rate results suggests that the 3- to 4-month old infants' competence in discriminating burst cues may be better assessed with a no-ITI cardiac paradigm than with an habituation-dishabituation cardiac procedure. Therefore, the present experiment employed a no-ITI paradigm similar to that used by Leavitt et al. to investigate further the burst discrimination of 3-4-month old infants. In this study 20 tokens of one syllable (for example, [bu]) were followed immediately by 20 tokens of a change syllable ([gu]) and discrimination was indexed by an OR to the stimulus shift.

METHOD

Subjects

Twelve infants, aged 3-4 months (mean = 3 mos, 1.5 wks), served as subjects. The participation of parents in the greater Madison area was solicited by a letter describing the research and a follow-up phone call. A total of 29 infants was tested with 14 (48 percent) eliminated on the basis of predetermined state criteria, 2 because of equipment problems, and 1 due to experimenter error.[1] The 12 remaining infants included 5 males and 7 females.

Apparatus

Each subject was tested in an infant seat positioned on a table-like platform in an Audio-Suttle sound-attenuated chamber. Throughout the session, the parents and experimenter were able to monitor visually the infant's behavior over a closed-circuit television system. Stimuli were played to the subject on a TEAC 3300S 2-track tape deck coupled to a Bogen Challenger amplifier and Hewlett-Packard attenuator. An Acoustic Research 2 ax speaker, located 40" in front of the infant, presented the stimuli at 70 ± 1 dB (A) SPL against a background level of 27 dB (A). Sound level measurements were made with a General Radio Sound Level Meter (#1551-C, microphone #1560-P5) placed at the site of the infant's head.

A stimulus artifact on the second channel of the stimulus tape, denoting stimulus onset and change, occurred coincident with the first and twenty-first stimuli of each 20/20 trial. A Scientific Prototype audio threshold relay detected the stimulus artifact and converted it into a suitable pulse for recording on one channel of a Sony TC 756 2-track tape deck. Cardiac activity was detected by Beckman biopotential miniature skin electrodes and amplified by a Gilson polygraph. The two active electrodes were placed 2-3 cm above the right nipple and approximately 2-3 cm above and to the left of the navel. A ground electrode was placed 2-3 cm above the left nipple. Sites for electrode placement were prepared with alcohol and the electrodes were attached with either Beckman or Beck-Lee paste and micropore tape. An adjustable pulser converted each R wave in the electrocardiogram (EKG) into a square pulse suitable for recording on the second channel of the Sony TC 756 tape deck.

---

[1] Of the 14 infants rejected for state, 10 had at least one behaviorally acceptable trial. Thus, only 4 infants (15 percent) did not contribute acceptable data for the first trial, suggesting that because of the low attrition rate this paradigm is a desirable one for infant researchers.

Stimuli

The natural speech stimuli [bu] and [gu], produced by an adult male speaker, were used in constructing the experimental [bu] and [gu] stimuli shown in Figure 2. The experimental [bu] was the natural [bu] initially produced by the speaker. The experimental [gu] was produced by removing the burst portion from the natural [bu] and replacing it with the burst portion of the natural [gu]. This experimental stimulus was consistently identified as a [gu] by over 50 adult Ss. The duration of the [bu] was 450 msec with an 8 msec burst, whereas the 474 msec [gu] contained a 32 msec burst. The construction and recording of the stimuli were carried out on the PCM system at Haskins Laboratories (Cooper and Mattingly, 1969). These were the identical stimuli employed in the Miller et al. (1975) study.

Procedure

Upon arrival at the laboratory, the parents were briefed on the procedures and purposes of the research and their consent was solicited prior to the test session. The infant was then placed in the infant seat in the testing chamber and the electrodes were affixed to the infant's chest. When the infant was judged to be in a quiet, alert state, stimulus presentations began. The general parameters of the 20/20 paradigm and the orders presented to subjects are depicted in Table 1 together with the contrasting features of the 8/2 procedure employed in the Miller et al. (1975) study. As can be seen in Table 1, each subject was presented with 4 20/20 sequences (ISI = 1 sec), each separated by a 30-second pause. The order of stimulus change within the four trials was alternated from [bu] → [gu] to [gu] → [bu] (or vice versa) and the order of presentation on the first trial was counterbalanced across subjects, such that half of the subjects received a [bu] → [gu] shift on trial 1 (Group A) and half a [gu] → [bu] shift (Group B). The duration of each 20/20 sequence was 1 minute and that of the entire experimental session approximately 5.5 minutes.

Throughout the session, an assistant seated inside the chamber and out of the infant's sight observed and recorded the infant's behavior using a closed-circuit TV monitor. Behavior recording occurred for 5 seconds prior to and 10 seconds following each trial onset and change and included visual behavior (for example, fixation, eye widening), body movements, vocalization, sucking behavior, and states of arousal (for example, fussy, drowsy, alert). Subjects were eliminated from the study only if they exhibited excessive fussiness, drowsiness, and/or large movements during the behavior recording periods (cf. Leavitt, 1975, for further details of recording and acceptance criteria).

Data Reduction

Each R wave of the infant's EKG was recorded as a square pulse on audio tape and the R-R (interbeat) intervals for 5 prestimulus and 15 poststimulus seconds were computed by a PDP-12 computer. These data were then converted by a Datacraft computer into a beats-per-minute (bpm) measure for each pre- and poststimulus second. Prestimulus level was calculated for one second prior to each trial onset and change. Analyses were performed on difference scores calculated by subtracting this prestimulus level from each of the subsequent

FIGURE 2

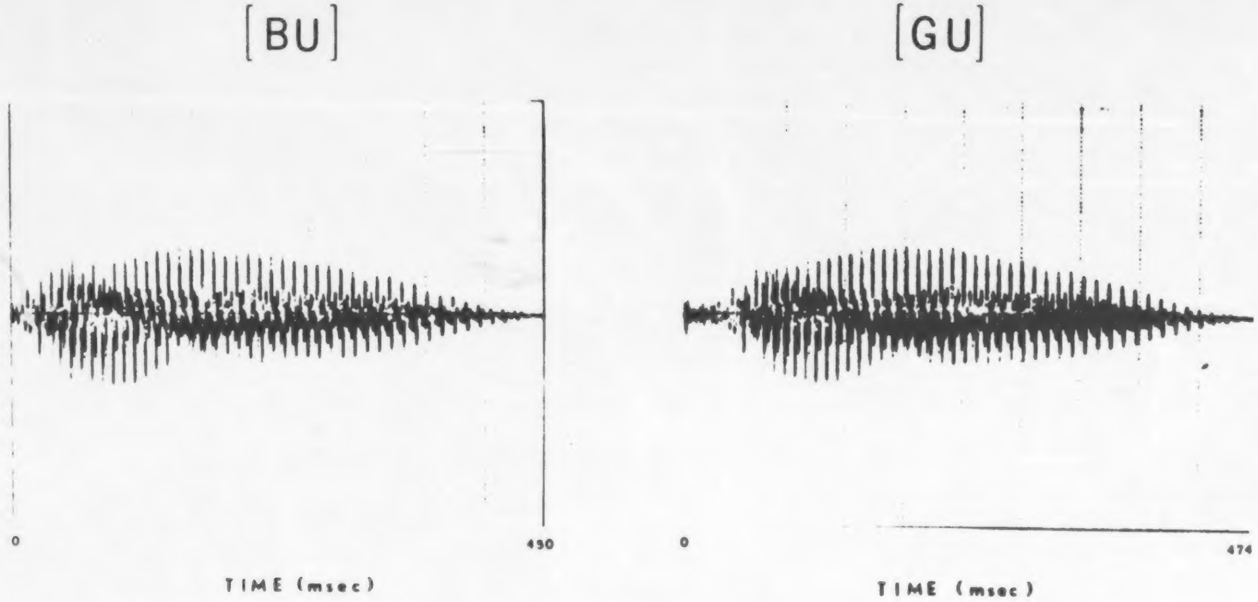


Figure 2: Oscillograms of the experimental [bu] and [gu] stimuli.

TABLE 1: General parameters of the 20/20 HR discrimination paradigm of the present study and the 8/2 habituation-dishabituation paradigm employed by Miller et al. (1975a).

	Parameters					Presentation Orders	
8/2 Paradigm	T1	T2	T8	T9	T10	A BU → GU	
(Miller et al., 1975a)	00000000	00000000 ...	00000000	11111111	11111111	B GU → BU	
	↑ ISI=1 sec	↖ ITI=25-35 secs	↗				
20/20 Paradigm	1 ↓		21 ↓				
(present study)	00000000000000000000		11111111111111111111			A 1 BU → GU	B 1 GU → BU
	↑ ISI=1 sec					2 GU → BU	2 BU → GU
						3 BU → GU	3 GU → BU
						4 GU → BU	4 BU → GU

15 poststimulus seconds. F-tests were performed by a Datacraft computer on the mean difference scores to determine significant departures from prestimulus level for each poststimulus second and trend analyses were carried out by a UNIVAC 1110 computer.[2]

RESULTS

Onset Data

The onset data of the 20/20 procedure are displayed for the two orders of presentation on trial 1 (as detailed in Table 1) in Figure 3. These data were subjected to an analysis of variance for trends with Order of Presentation on trial 1 (A = [bu] → [gu] vs B = [gu] → [bu]) as a between-subjects factor and within-subjects factors of Shift Condition within the 4 trials ([bu] → [gu] vs [gu] → [bu]), Trial Blocks (TB 1 = trials 1 and 2 vs TB 2 = trials 3 and 4), and Seconds (15).

A reliable orienting response to onset over all trials was shown in the significant quadratic trend over Seconds, $F(1,10) = 42.46$, $p < .001$. Furthermore, a significant main effect for Trial Blocks, $F(1,10) = 6.6$, $p < .01$, suggested that the initial OR habituated from the first to the second half of the session. Although no significant main effect for Order of Presentation was obtained, a significant quadratic trend over Seconds x Order x Condition interaction, $F(1,10) = 23.07$, $p < .001$, was observed. In addition, the quadratic trend over Seconds x Condition x Trial Blocks interaction, $F(1,10) = 7.75$, $p < .025$, was also found to be reliable. As can be seen in Figure 3, these interactions indicate that the magnitude of the initial orienting response varied in the two orders of initial presentation as a function of stimulus shift and the first vs. the second half of the session.

Change Data

The change data are separated for the two orders, A and B, in Figure 4. F-tests on the difference scores of these data revealed that orienting occurred on every trial in which there was a [bu] → [gu] shift (trial 1, $p < .01$; trial 2, $p < .05$; trial 3, $p < .01$; trial 4, $p < .050$.[3] In contrast, no significant orienting (or acceleration) occurred on those trials in which there was a [gu] → [bu] shift. A 4-way analysis of variance identical to that employed for the onset data, was performed on the change data and confirmed the existence of differential responding to the stimulus change. A significant main effect of Shift Condition, $F(1,10) = 6.05$, $p < .05$, a significant Seconds x Shift Condition interaction, $F(14,140) = 2.75$, $p < .005$, and a significant cubic trend over Seconds x Shift Condition interaction,

---

[2] Wilson's (1974) CARDIVAR package was employed in preparing the R-R interval data for subsequent analyses. The F-test and trend analysis programs were developed and generously made available to us by Dr. F. K. Graham.

[3] All significance levels for these F-test analyses were converted to the Bonferroni t (Myers, 1972).

FIGURE 3

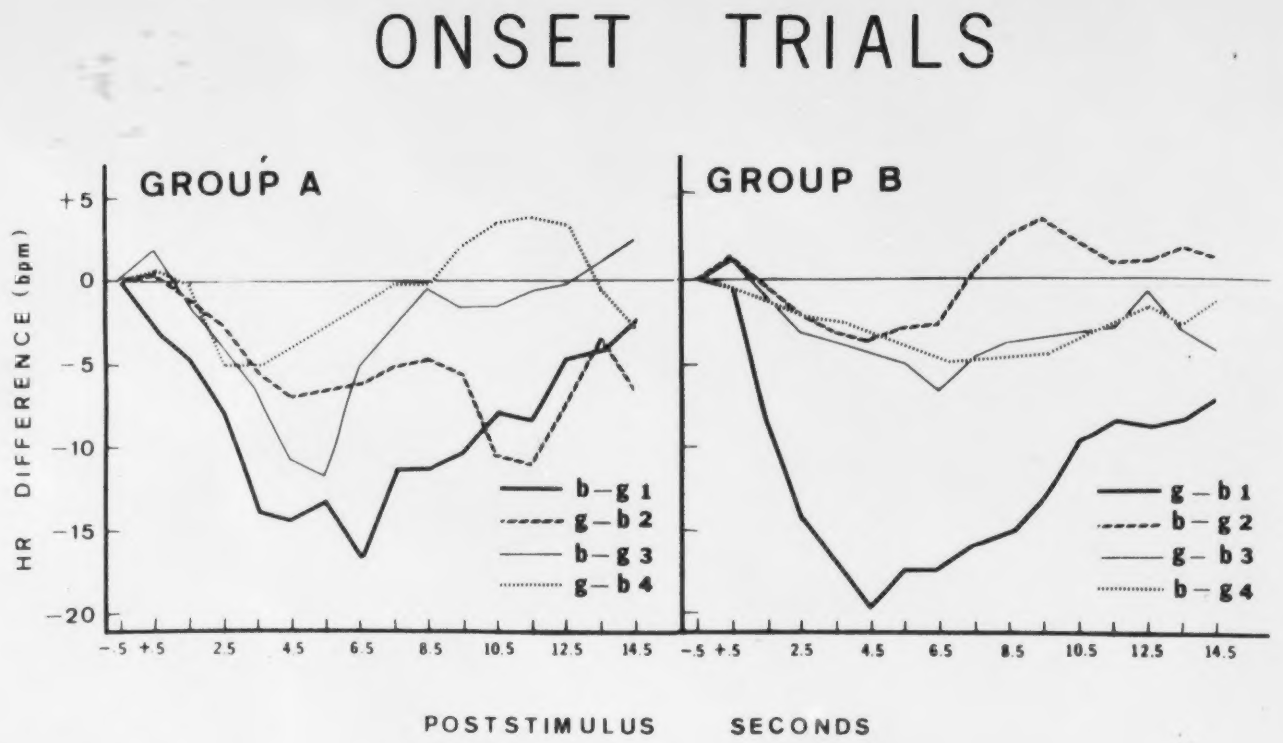


Figure 3: Onset data for the 4 20/20 trials, grouped according to initial shift condition. (A = [bu] → [gu] on trial 1 vs.B =[gu] → [bu] on trial 1)

FIGURE 4

CHANGE TRIALS

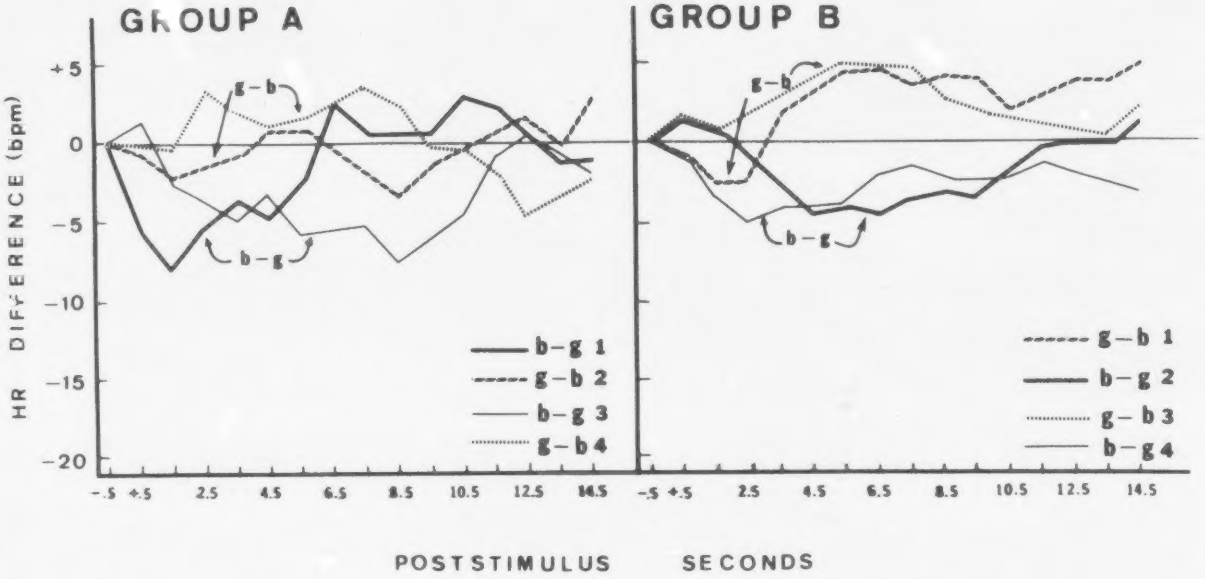


Figure 4: Change data for the 4 20/20 trials, grouped according to initial shift condition. (A = [bu] → [gu] on trial 1 vs. B = [gu] → [bu] on trial 1). Broken lines = [gu] → [bu] shift trials; solid lines = [bu] → [gu] shift trials.

$F(1,10) = 19.25$, $p < .005$, all indicated that there were different responses to the two types of stimulus change. The absence of any order effects confirmed the consistency of the stimulus shift effects across both orders A and B.[4]

DISCUSSION

Asymmetry in Discrimination

Two possible interpretations that may be offered for the asymmetry observed in this study are related to the construction of the experimental stimuli. The first possibility is that the [gu] stimulus may have been more salient to the infant, and therefore a greater elicitor of orienting. The [gu] burst, relative to the [bu] burst, is slightly longer in duration, has a greater energy concentration within a more restricted frequency range, and is acoustically incongruous with the [bu] formant transitions which follow. Any or all of these features may have enhanced the saliency of the [gu] stimulus.[5]

A second possibility is that during the first 20 stimuli, the burst cue is being adapted in a manner similar to that reported in adults for a variety of other speech cues, including bursts (Eimas and Corbit, 1973; Ades, 1974; Cooper, 1974; Blumstein and Stevens, 1975; Diehl, 1975; Ganong, 1975; Tartter and Eimas, 1975; Morse, Kass and Turkienicz, 1976). In the present study, if the [gu] burst is adapted, the remaining cues of the experimental [gu] stimulus are the formant transitions of the [bu]. Consequently, if adaptation occurs within the first 20 stimuli, the [gu] stimulus would become a [bu], and hence the shift to [bu] might not be discriminable. In contrast, when the [bu] burst adapts, a [bu] remains and would be discriminated from a shift to [gu]. Although adult studies of speech adaptation have not attempted to determine whether reliable adaptation effects can occur within only 20 presentations of a stimulus, this possibility cannot be ruled out, either in the adult or in the infant.

Although the asymmetry observed in the present study was not anticipated, asymmetries in infant speech discrimination have also been reported elsewhere (Butterfield and Cairns, 1974). Since Eimas (1975b) has suggested that the process of adaptation may be responsible for much of the evidence of infant speech discrimination obtained with the nonnutritive sucking paradigm, more direct tests of this adaptation account using infant paradigms with adult listeners may greatly aid in elucidating the mechanisms responsible for these asymmetries

---

[4] Subsequently, a group of 12 infants was tested in a control condition (no stimulus change), that included one 20/20 trial of either [bu] or [gu]. F-tests and trend analyses performed on these data revealed no signifiant cardiac deceleration following the twenty-first stimulus. These data confirm the conclusion that the OR recovery to the [bu] → [gu] shifts actually reflected burst discrimination, rather than cyclic changes in cardiac activity.

[5] Although inspection of Figure 3 suggests that greater onset orienting did occur to the [gu] stimulus, there is no statistical support for this observation. However, a similar study of infant burst discrimination by Miller, Goy, Morse and Dorman (1975) did find statistical evidence of greater initial orienting to [gi] and [bi], using burst stimuli constructed in a manner similar to those of the present study.

Discrimination: 8/2 vs. 20/20 Paradigm

The major implication of the present experiment is that infant burst discrimination is dependent upon the particular paradigm employed. These results revealed that infants are capable of this discrimination when tested with the 20/20 paradigm, yet do not appear so within the 8/2 paradigm (Miller et al., 1975a). Since no study has reported cardiac dishabituation to auditory stimuli in infants younger than 4 months, it is possible that the lack of evidenced discrimination in the Miller et al. (1975a) study may reflect an inability of young infants to demonstrate OR dishabituation to any change in auditory stimulation. This interpretation is consistent with the data reported by Brown et al. (1975, in press) that suggest developmental trends in the characteristic properties of orienting behavior (that is, initial OR, habituation, dishabituation) to auditory stimuli. However, since Adkinson and Berg (1974) have observed cardiac dishabituation to visual stimuli in newborns, this conclusion remains somewhat tenuous.

Perhaps the more productive way of interpreting the difference between these two studies would be to examine the physical parameters of the two paradigms employed (cf. Table 1). Two obvious parametric differences resulting in the different stimulus distributions of these two paradigms are: 1) the number of familiar stimuli preceding the stimulus shift, and 2) the ITI's separating blocks in the 8/2 procedure, which are absent in the 20/20 paradigm. Since there are actually fewer tokens of the familiar stimulus presented to the infant prior to the change in the 20/20 paradigm (20, as opposed to 64 in the 8/2 procedure), any differences in memory for the familiar syllable cannot be due to the total number of prechange exemplars.

If, instead, the ITI's of the 8/2 paradigm were primarily responsible for the different results obtained in these two experiments, then it may be because these lengthy silent intervals in some manner imposed too great a burden upon the infant's processing of these burst stimuli. In other words, the distribution over time of the stimuli in the 8/2 procedure may have resulted in less consolidation of the stimulus being stored in memory and/or some decay during the ITI of the "neuronal model" (Sokolov, 1963) of the burst stimulus. Consequently, the habituation observed in the 8/2 paradigm may have reflected the development of a more general model of the stimulus presented, and thus the absence of discrimination in this paradigm is not surprising. Unfortunately, the development of a parametric research program to answer these questions may be complicated by several factors. First, the general parameters of the habituation-dishabituation paradigm were originally conceptualized to include recovery time between trials for an OR to stimulus offset. In addition, some recent work (Roth and Morse, 1975) suggests that for speech sounds, an infant's orienting response to initial stimulus onset may require some 20-30 seconds for recovery. Thus, one cannot simply vary the ITI between blocks of stimuli in moving from a 20/20 paradigm to an 8/2 procedure. Although the results of the present study do not resolve these questions about the processes underlying cardiac measures of infant discrimination, they do demonstrate that with one cardiac paradigm, burst discrimination does occur in early infancy.

In conclusion, the results of the present experiment have several implications for our understanding of the development of infant speech perception. First, they reveal that young infants can discriminate very brief burst cues in stop consonants, thus adding to the list of important acoustic events in the speech signal to which infants are sensitive at a very early age. Second, the consistent pattern of asymmetry suggests that more direct tests of adaptation proposals for infant (and adult) burst discrimination may greatly enhance our understanding of the mechanisms that underlie infant speech perception. Third, the burst discrimination obtained with the 20/20 procedure of the present study suggests (as Leavitt et al. observed) that a no-ITI paradigm may be more useful in studying the speech discrimination of infants younger than 4 months than the more traditional habituation-dishabituation procedure. The recent evidence of categorical discrimination for place of articulation using the 20/20 paradigm in 3- to 4-month old infants (Miller and Morse, 1976) further underscores the usefulnesss of this paradigm in studying infant speech discrimination.

REFERENCES

Ades, A. (1974) A bilateral component in speech perception. J. Acoust. Soc. Am. 56, 610-616.

Adkinson, C. D. and Berg, W. K. (1974) Habituation and dishabituation of cardiac orienting in newborns. Psychophysiol. 11, 219.

Berg, W. K. (1972) Habituation and dishabituation of cardiac responses in 4-month-old, alert infants. J. Exp. Child Psychol. 14, 92-107.

Berg, W. K. (1974) Cardiac orienting responses of 6- and 16-week-old infants. J. Exp. Child Psychol. 17, 303-312.

Blumstein, S. E. and Stevens, K. N. (1975) Property detectors for bursts and transitions in speech perception. J. Acoust. Soc. Am. 57 (S)1, 52.

Brown, J. W., Leavitt, L. A. and Graham, F. K. (1975) Infant cardiac response during a six-week transition period. Research Status Report 1, Infant Development Laboratory. (Madison: University of Wisconsin), 65-89.

Brown, J. W., Leavitt, L. A. and Graham, F. K. (in press) Response to auditory stimuli in six- and nine-week-old infants. Developmen. Psychobiol.

Butterfield, E. and Cairns, G. (1974) Discussion summary - Infant perception research. In Language Perspectives - Acquisition, Retardation, and Intervention, ed. by R. Schiefelbusch and L. Lloyd. (Baltimore, Md.: University Park Press), pp. 75-102.

Cole, R. A. and Scott, B. (1974a) The phantom in the phoneme: Invariant characteristics of stop consonants. Percept. Psychophys. 15, 101-107.

Cole, R. A. and Scott, B. (1974b) Toward a theory of speech perception. Psychol. Rev. 81, 348-374.

Cooper, F. S. and Mattingly, I. G. (1969) Computer-controlled PCM system for investigation of dichotic speech perception. Haskins Laboratories Status Report on Speech Research SR-17/18, 17-21.

Cooper, W. E. (1974) Adaptation of phonetic feature analyzers for place of articulation. J. Acoust. Soc. Am. 56, 617-627.

Diehl, R. L. (1975) The effect of selective adaptation on the identification of speech sounds. *Percept. Psychophys.* *17*, 48-52.

Dorman, M. F., Studdert-Kennedy, M. and Raphael, L. J. (in press) Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Percept. Psychophys.*

Eimas, P. D. (1974) Auditory and linguistic processing of cues for place of articulation by infants. *Percept. Psychophys.* *16*, 513-521.

Eimas, P. (1975a) Auditory and phonetic coding of the cues for speech: discrimination of the [r-l] distinction by young infants. *Percept. Psychophys.* *18*, 341-347.

Eimas, P. D. (1975b) Speech perception in early infancy. In *Infant Perception*, ed. by L. B. Cohen and P. Salapatek. (New York: Academic Press).

Eimas, P. D. and Corbit, J. D. (1973) Selective adaptation of linguistic feature detectors. *Cog. Psychol.* *4*, 99-109.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. and Vigorito, J. (1971) Speech perception in infants. *Science* *171*, 303-306.

Fischer-Jørgenson, E. (1972) Perceptual studies of Danish stop consonants. Annual Report VI. (Copenhagen: Institute of Phonetics, University of Copenhagen).

Ganong, W. F. (1975) Phonetic adaptation measured on two continua between the same syllables. *J. Acoust. Soc. Am.* *58*, (S)1, 58.

Lasky, R. E., Syrdal-Lasky, A. and Klein, R. F. (1975) VOT discrimination by four to six and a half month old infants from Spanish environments. *J. Exp. Child Psychol.* *20*, 215-225.

Leavitt, L. A. (1975) State rating scales used in the Infant Development Laboratory. *Research Status Report* *1*, Infant Development Laboratory. (Madison: University of Wisconsin), 341-345.

Leavitt, L. A., Brown, J. W., Morse, P. A. and Graham, F. K. (in press) Cardiac orienting and auditory discrimination in six-week infants. *Developmental Psychology*.

Liberman, A. M. (1970) The grammars of speech and language. *Cog. Psychol.* *1*, 301-323.

Liberman, A. M., Cooper, F. S., Shankweiler, D. and Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychol. Rev.* *74*, 431-461.

Liberman, A. M., Delattre, P. C. and Cooper, F. S. (1952) The role of selected stimulus variables in the perception of unvoiced stop consonants. *Am. J. Psychol.* *65*, 497-516.

Miller, C. L., Goy, E. R., Morse, P. A. and Dorman, M. F. (1975) Selected problems in infant burst discrimination. *Research Status Report* *1*, Infant Development Laboratory. (Madison: University of Wisconsin), 171-181.

Miller, C. L. and Morse, P. A. (1976) The "heart" of categorical speech discrimination in young infants. *J. Sp. Hear. Res.* *19*, 578-589.

Miller, C. L., Morse, P. A. and Dorman, M. F. (1975) Memory for burst cues in infant burst discrimination. *Research Status Report* *1*, Infant Development Laboratory. (Madison: University of Wisconsin), 149-169.

Miller, J. L. (1974) Phonetic determination of infant speech perception. Unpublished doctoral dissertation, University of Minnesota, Minneapolis, Minnesota.

Moffitt, A. R. (1971) Consonant cue perception by twenty-to-twenty-four-week-old infants. *Child. Dev.* *42*, 717-731.

Morse, P. A., Kass, J. E. and Turkienicz, R. (1976) Selective adaptation of vowels. Percept. Psychophys. 19, 137-143.

Morse, P. A., Leavitt, L. A., Miller, C. L. and Romero, R. C. (in press) Overt and covert aspects of adult speech perception. J. Sp. Hear. Res.

Myers, J. A. (1972) Fundamentals of Experimental Design. (Boston: Allyn & Bacon, Inc.).

Roth, P. L. and Morse, P. A. (1975) An investigation of infant VOT discrimination using the cardiac OR. Research Status Report 1, Infant Development Laboratory. (Madison: University of Wisconsin), 207-218.

Sokolov, E. N. (1963) Perception and the Conditioned Reflex. (New York: Macmillan).

Swoboda, P. J., Morse, P. A. and Leavitt, L. A. (1976) Continuous vowel discrimination in normal and at-risk infants. Child Develop. 47, 459-465.

Tartter, V. S. and Eimas, P. D. (1975) The role of auditory feature detectors in perception of speech. Percept. Psychophys. 18, 293-298.

Till, J. (1976) Infant's discrimination of speech and nonspeech stimuli. Unpublished doctoral dissertation, University of Iowa, Iowa City, Iowa.

Trehub, S. E. and Rabinovitch, M. S. (1972) Auditory-linguistic sensitivy in early infancy. Developmen. Psychol. 6, 74-77.

Wilson, R. A. (1974) CARDIVAR: The statistical analysis of heart rate data. Psychophysiol. 11, 76-85.

Cinefluorographic and Electromyographic Studies of Articulatory Organization*

Thomas Gay†

ABSTRACT

Articulatory studies dealing with the spreading effect of coarticulation and the control of speech rate are reviewed. The results of some more recent research is then reviewed with a view toward refining earlier formulations. The suggestion is made, in contrast to traditional views, that the segmental input to the speech string is governed by simple rules that operate within a limited coarticulatory field, while the temporal organization of the string requires complex articulatory adjustments based on advanced information obtained from a higher level scan-ahead mechanism.

INTRODUCTION

This paper summarizes the results of several experiments that used the techniques of cinefluorography and electromyography to study the organization of speech gestures. As such, it does not represent a comprehensive review of current speech production theory, but rather is directed towards a discussion of several specific issues that are best studied by these techniques: the dynamics of articulatory movements and the motor command structure that underlies those movements.

Although speech is usually described in terms of a string of invariant segmental units (phonemes), the act of speaking imposes on this string a complex encoding. This is a consequence of the series of events that comprise the speech production chain: the conversions of motor command-to-muscle contraction, muscle contraction-to-vocal tract shape, and vocal tract shape-to-acoustic signal. The result of this encoding is observed as variation in both the production of a given phone and in its acoustic representation. This paper will be concerned with allophonic variation as it appears at the articulatory level, specifically, variations that arise from changes in

---

*This paper was presented at the U.S. Japan Joint Seminar on Dynamic Aspects of Speech Production, Hill-Top Hotel, Tokyo, Japan, 7-10 December, 1976.

†Also University of Connecticut Health Center, Farmington.

Acknowledgment: The research reported here was supported by grants from The National Institute of Dental Research (DE-01774), National Institute of Neurological and Communicative Diseases and Stroke (NS-10424) and The National Science Foundation (GSOC 740 3725).

phonetic context, and variations that arise from changes in the suprasegmental structure of the string, particularly changes in speech rate.

THE ORGANIZATION OF SEGMENTAL GESTURES

Coarticulation is usually defined as allophonic variation of a given phone due to changes in its phonetic environment. The production of a phone can be conditioned by a phone that either precedes it (left-to-right or carryover effects) or follows it (right-to-left or anticipatory effects).

Anticipatory coarticulation effects are essentially timing effects: movements toward some parts of a feature target of a given segment begin before others. Kozhevnikov and Chistovich (1965) studied the anticipation of lip rounding that occurs when a rounded vowel follows a consonant and suggested that the forward extent of this anticipatory gesture was the first consonant in the sequence. Daniloff and Moll (1968) also showed that lip rounding can begin ahead of the syllable boundary and across as many as four consonants preceding the vowel. In their experiment, anticipation of lip rounding for the vowel /u/ was studied for a number of mono- and disyllabic single and two-word utterances imbedded in sentence frames, using lateral view x-ray motion pictures. Onset of lip rounding usually begins with the first consonant of the utterance. Another type of anticipatory coarticulation was shown to exist by Öhman (1966). In a spectrographic study of coarticulation in VCV sequences, Öhman concluded that the variability observed in transitional movements to the consonant could be predicted by the formant frequencies of the second vowel. This led Öhman to conclude further that vowel-to-vowel movement in a VCV is essentially diphthongal with the consonant simply superimposed on the basic gesture; in other words, anticipatory movements toward the second vowel begin independently from those toward the consonant. These studies, among others, suggest that articulatory encoding is a complex phenomenon whose effects can spread across several adjacent segments. Most support, either explicitly or implicitly, Henke's (1966) articulatory model that proposes the operation of a mechanism that scans future segmental inputs, or features thereof, and sends commands for the immediate attainment of those feature targets that would not interfere with the attainment of immediately intervening articulations. However, in two recent studies, both cinefluorographic (see Gay, in press) and electromyographic (see Gay, 1974b), evidence was used to argue against the pervasiveness of anticipatory coarticulation in speech.

In the cinefluorographic experiment, conventional high speed (60 fps) lateral view x-ray films were recorded from two subjects who produced various VCV syllables that contained the vowels /i,a,u/ and the consonants /p,t,k/ in all possible combinations. Articulatory movements were tracked by recording the positions, frame-by-frame, of 2.5 mm diameter lead pellets that were attached to the upper and lower lips, jaw, and several locations along the surface of the tongue relative to a reference pellet attached at the embrasure of the upper central incisors. These data will be used to explore the question of whether, in a VCV utterance, an intervening consonant constrains movements of the articulators, in particular the tongue body and lips, from one vowel to another; is the movement from one vowel to another essentially diphthongal or is it somehow locked to the consonant (Öhman's hypothesis), and does the lip rounding gesture for the postvocalic rounded vowel begin ahead of the intervocalic consonant? (Henke's model).

The dynamic properties of articulatory movements in a VCV sequence are illustrated in Figure 1 for an utterance where the intervocalic consonant is /p/. This figure shows the movement tracks of the tongue body, lips, and jaw in the height dimension for the sequence /ipa/. Each track was graphed from discrete points measured every film frame, that is, at approximately 17 msec intervals. Measurements begin during the closure period of the initial /k/ and end at the time of closure for the final /p/; 0 on the abscissa corresponds to the time of intervocalic consonant closure. This figure illustrates the general finding that the intervocalic consonant affects the timing of the movements of the tongue body from vowel to vowel. The movement of the tongue body from the first vowel to the second vowel does not begin until after closure for the intervocalic consonant is completed. This was found to be a salient feature in the production of all VCV utterances. Consonant constraints on vowel-to-vowel movements were as evident in the front-back dimension as in the height dimension, and the rules that apply to /p/ also apply when the intervocalic consonant is either /t/ or /k/. The only variability in the timing effect appears in the delay time between consonant closure and tongue body movement. While the lag was usually of the order of 30 msec, it varied anywhere from 10-60 msec. This figure also shows that the movements of the tongue body, because they begin ahead of those for the jaw, are probably independent from jaw movements towards the vowel. As is also evident in this figure, upper lip contributions to lip closure were negligible. Finally, this subject showed a pattern of lip closure that was often characterized by continued compression throughout the closure period.

Perhaps the best illustration of consonantal constraints on tongue body movements is one where the first and second vowels of the utterance are the same. Figure 2 shows the movement tracks for the jaw and four tongue pellets during the production of /iti/ for Subject FSC. Instead of the tongue maintaining the /i/ target during the consonant, the tongue blade and both tongue body pellets show movement throughout the consonant gesture. The blade and anterior tongue body pellet appear to shadow movements of the tip while the posterior tongue body pellet moves in the opposite direction (lower), probably in a facilitory gesture. However, since the tongue body gesture (when it does appear) does not reach a specific, repeatable location, it is not interpreted as being target directed.

The discontinuity of the vowel target in an utterance where the same two vowels are separated by a consonant is also evident at the EMG level (see Gay, 1974b). The average EMG activity of the genioglossus muscle for the sequences /ipi/ and /iti/ is illustrated in Figure 3. The genioglossus muscle, which comprises the bulk of the tongue body, is primarily responsible for the protruding and bunching associated with the vowel /i/. This figure shows three separate peaks associated with the utterance. The first peak corresponds to the initial /k/, while the second and third correspond to the first and second vowels. Of particular interest is the deep trough that separates the two vowel peaks. The presence of a trough, which signifies a cessation of genioglossus activity, suggests that the two vowels, although phonetically identical, are organized as two separate events. If the movement of the tongue body during the production of the consonant as observed in the x-ray data (Figure 2) was due solely to the movement of other articulatory

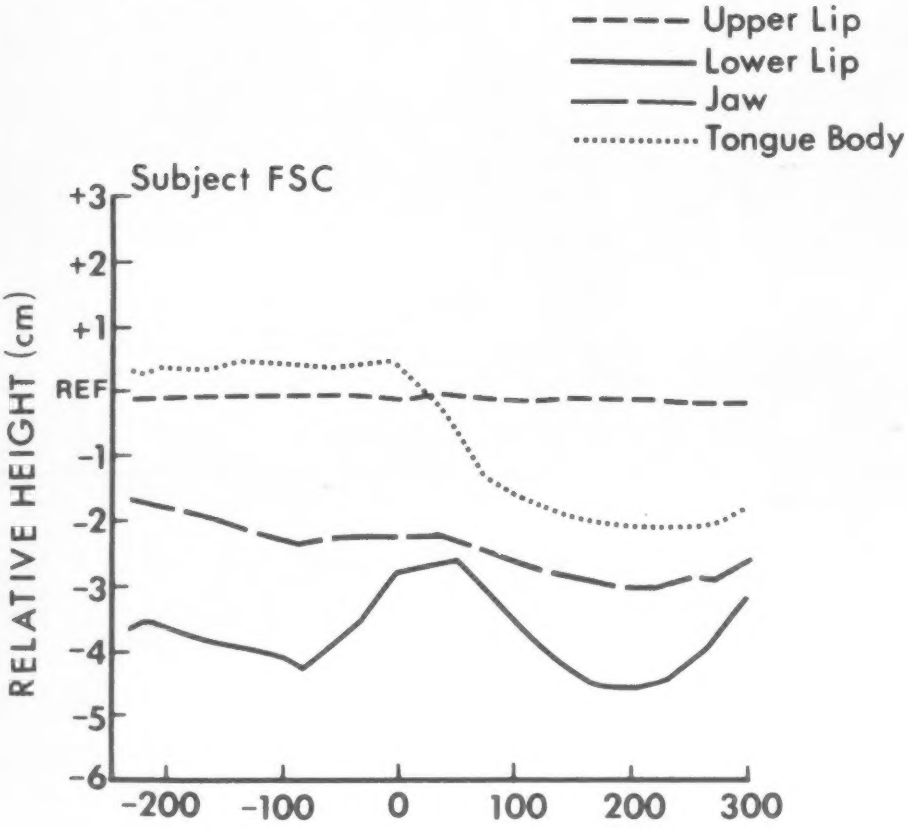


Figure 1: Movement tracks for the sequence /ipa/. Relative tongue height (ordinate) is plotted as a function of time (abscissa) in msec.

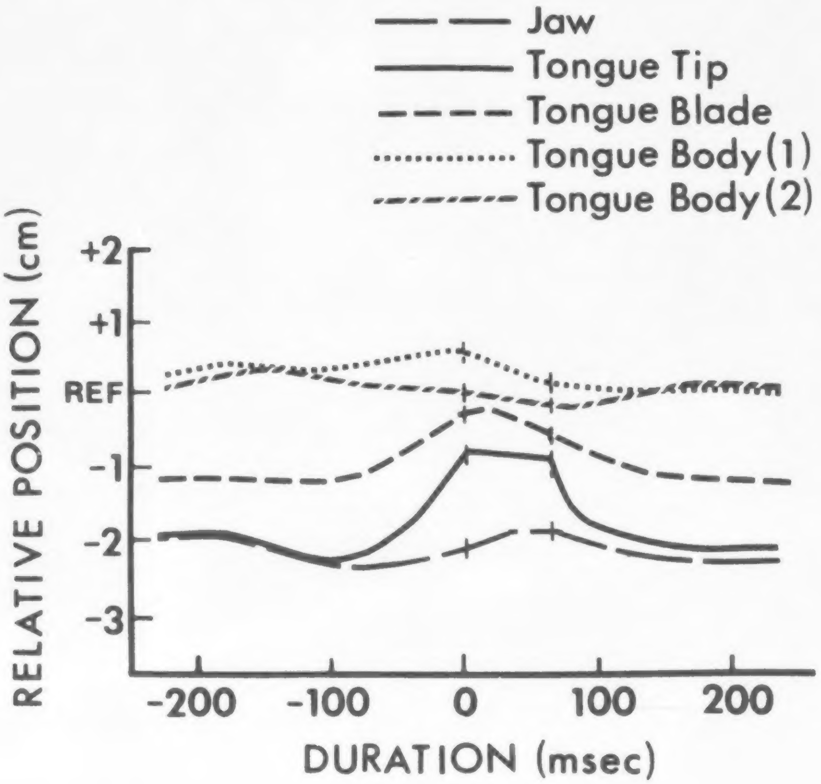


Figure 2: Movement tracks for the sequence /iti/ Subject FSC.

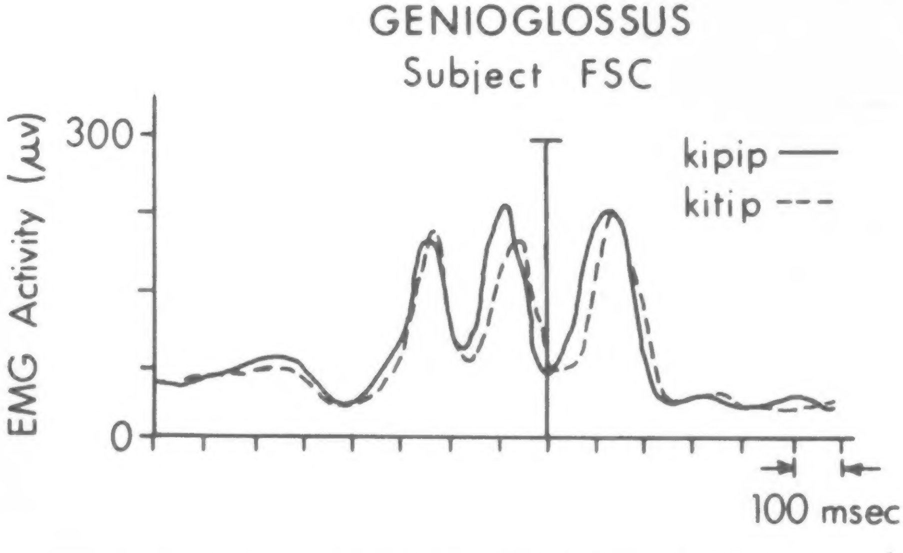


Figure 3: Average genioglossus muscle activity for the utterances /ipi/ and /iti/, Subject FSC.

structures, such as the tongue tip or jaw, positional constancy would exist at the EMG level in the form of one broad genioglossus peak across the entire utterance. However, the existence of two distinct peaks separated by a deep trough suggests that the intervocalic consonant has more than a passive effect on tongue body movement.

For VCV utterances containing either /p/, /t/ or /k/ as the intervocalic consonant, the usual sequence of articulatory events is as follows: movements of the jaw, tongue body and primary articulator begin at about the same time, with jaw closing continuing past the time of occlusion for the consonant. Shortly after closure for the consonant occurs, tongue body movement toward the second vowel begins. This movement is followed independently, by jaw opening and release of the consonant. Articulatory movements for the postvocalic vowel always begin between the time of consonant closure and consonant release. Constraints of the intervocalic consonant are also evident at the EMG level in the form of a separate muscle peak for each syllable.

The data of both the cinefluorographic and electromyographic experiments, in showing consonant constraints on vowel movement in a VCV utterance, argue against Öhman's (see Öhman, 1966) hypothesis that suggests vowel-to-vowel movement is essentially diphthongal. If Öhman's model were correct, tongue body movements toward the second vowel would begin at about the time of onset of closing for the consonant. However, movement toward the second vowel begins much later, some 10-60 msec after closure for the consonant has already been completed. This suggests that either the tongue body itself attains a target during consonant production, or more likely, that the intervocalic consonant and the following vowel are linked in a basic gesture. The very short lag time between consonant closure and movements toward the second vowel suggest the latter possibility.

In addition to placing constraints on the movements of the tongue body from one vowel to another, an intervocalic consonant also affects the onset of lip rounding for a rounded second vowel. These constraints are evident at both the articulatory and EMG levels. Lateral view x-rays can provide an indication of lip rounding in the form of degree of lip protrusion. In those cases where a rounded vowel appears in a post-consonantal position, the rounding gesture, like tongue body movements, does not begin until after closure for the intervocalic consonant is completed. This is true even for the most sensitive case, namely, two rounded vowels separated by a close consonant. Figure 4 shows the movement tracks of lower lip height, lower lip protrusion, and tongue tip height plotted against the same baseline for the syllable /utu/, produced by Subject GNS. Even in this example, it is evident that the rounding feature of the first vowel is not continuous through the consonant. Rather, what appears to be an additional, although small, closing and protruding gesture is superimposed on the rounding pattern. This discontinuity of rounding during the consonant is also evident in the EMG data (Figure 5) that shows a trough in orbicularis oris muscle activity during the production of the same syllable. Both sets of data argue against the Daniloff and Moll (1968) anticipatory effect. An alternative interpretation of Daniloff and Moll's result is that the early onset of lip rounding corresponded to a closing or protruding gesture of one or more of the intervening consonants in the utterance (for example, the /n/, /s/, /t/ or /r/ in the word "construe") or some special property of the cluster itself. This explanation

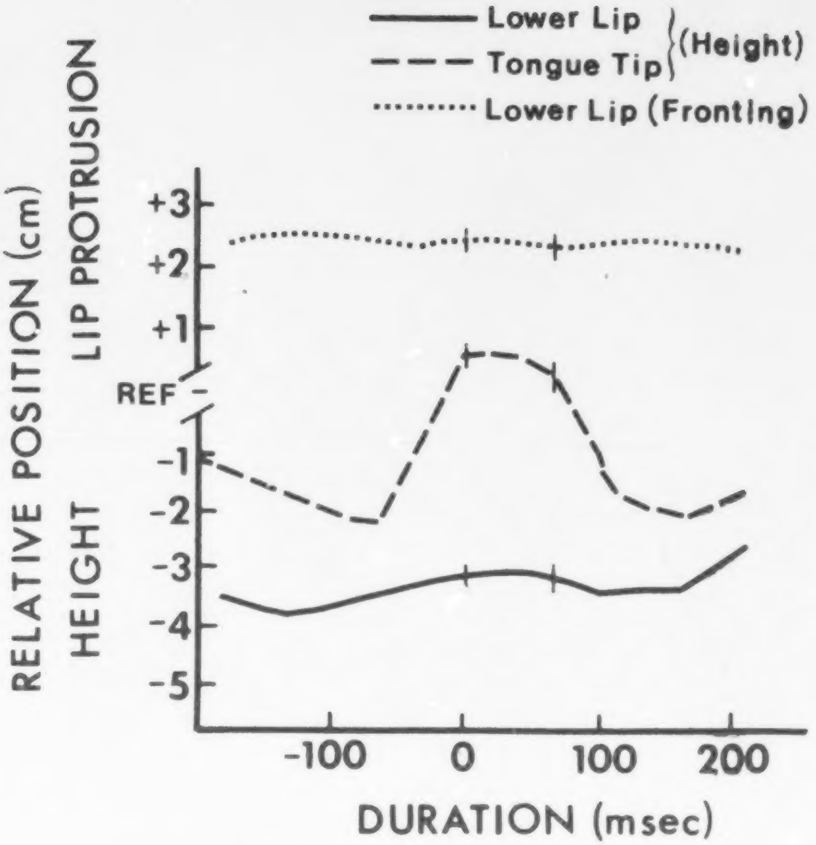


Figure 4: Movement tracks (tongue height and lip protrusion) for the utterance /utu/.

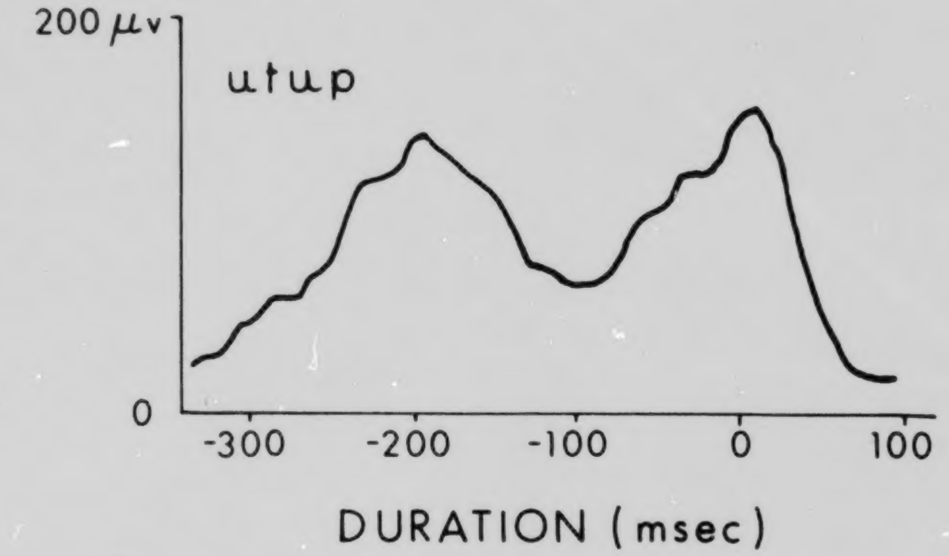


Figure 5: Average orbicularis oris muscle activity for the utterance /utu/.

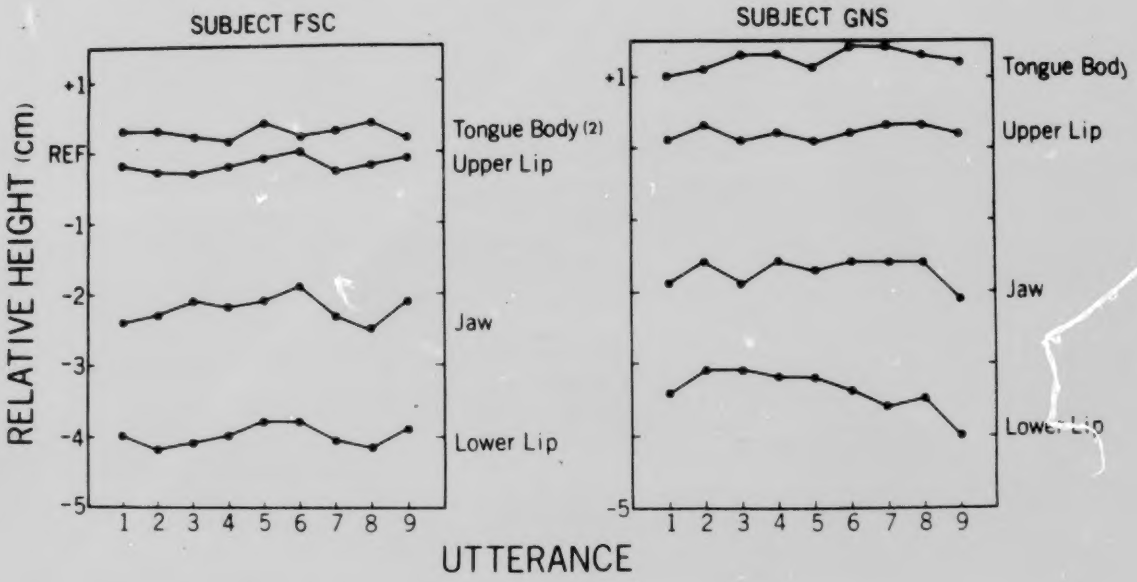


Figure 6: Coordinate positions (height) of upper lip, lower lip, jaw, and tongue body corresponding to the target positions of the vowel /i/.

seems even more likely in light of Bell-Berti and Harris' (see Bell-Berti and Harris, 1976) EMG data, which shows that the beginning of orbicularis oris muscle activity in utterances containing the syllables /stru/ and /stri/ occurs at the same time.

To summarize the data thus far: the relative timing of articulatory movements in a VCV sequence seems to be affected by the intervocalic consonant, even if the gesture for the consonant is not a contradictory one. The intervocalic consonant shows effects on tongue body movements toward and the lip rounding gesture for the second vowel at both the articulatory and EMG levels. Anticipatory movements toward the second vowel begin during the closure period of the intervocalic consonant, suggesting that the CV component of the VCV sequence is produced as a basic unit.

Unlike anticipatory coarticulation effects that are essentially timing effects, carryover coarticulation effects are usually considered as mechanical effects and exist in the form of variability in target (or target feature) positions as a function of changes in phonetic context. Although carryover effects have been shown to exist at both the EMG and articulatory levels, the pervasiveness of these effects is somewhat in doubt. In a study of the production of thirty-six CVC monosyllables, MacNeilage and DeClerk (1969) found that some aspect of the production of every phone was always influenced by a following phone. In particular, the size of the EMG signal would be different depending on the identity of the adjacent vowel or consonant. In countering the argument that a motor command representation of the phone shows less variability than an articulatory target representation, MacNeilage (1970) later proposed that the observed EMG variability reflected a complex motor strategy, the underlying goal of which is a relatively invariant articulatory end. The concept of an articulatory based target system as proposed by MacNeilage was further supported, at least for vowels by the cinefluorographic data of Gay et al. (Gay, Ushijima, Hirose and Cooper, 1974) and Gay (1974a). In the latter study, lateral view x-ray motion pictures were obtained from two speakers who produced the vowels /i,a,u/ in a variety of VCV contexts. The results of this experiment showed that for both subjects, the target positions for both /i/ and /u/, in both pre- and post-consonantal positions, remained quite stable (within 2-3 mm) across changes in the consonant and trans-consonantal vowel. Although target stability for /a/ was also the rule rather than the exception, some individual differences did appear. However, the articulatory variability, when it did appear, did not correlate with any acoustic variability.

Similar results were also reported in a more recent cinefluorographic study (Gay, in press). Carryover effects of an intervocalic consonant on the following vowel again appeared only for the open vowel /a/, and were reflected only in differences in jaw, and consequently, tongue body height. However, carryover effects of a preceding consonant on the attainment of target positions for the vowels /i/ and /u/ were minimal, with the tongue body targets of both vowels falling within a range of 2.5 mm for one subject and 3 mm for the other. This lack of variability is illustrated in Figure 6. The figure shows the relative positions of the upper lip, lower lip, jaw and tongue body at the time the tongue body reached its target (point of maximum displacement) for each of nine utterances containing the vowel /i/ in final position. As is evident, variability of tongue body target positions is minimal. Lower lip and jaw positions, on the other hand, vary within a larger range, approximately 5 mm for subject FSC and 10 mm for subject GNS. Interestingly, lower lip and jaw targets seem to vary independently from tongue body positions but covary for both subjects. This finding contradicts that of Hughes and Abbs (1976) who showed that mouth opening for /i/ remained relatively constant because of trade-offs between lower lip and jaw displacements. This type of equivalence was not evident in the present data for either /i/ or /u/.

Carryover effects, then, when they do appear, are unlike anticipatory effects in that they depend on the phonetic identity of the particular segment. Like anticipatory effects, however, carryover effects seem to spread no farther than the neighboring phone. Stability of tongue body targets for vowels (at least /i/ and /u/) is the rule rather than the exception. The only substantial articulatory variability occurred in jaw displacement, with /a/ showing the greatest effects and /u/ the least. However, variability in jaw displacement for /a/, as measured anteriorly at the incisors, might be either exaggerated or irrelevant in relation to variability that might exist in the pharyngeal constriction for /a/. Likewise, the variability of maximum jaw displacement for both /i/ and /u/ is unrelated to the variability observed in the position of the tongue body for those vowels. Thus, the two features, tongue body height and jaw displacement, are probably independent ones, with jaw opening being a facilitory gesture and an unmarked phonetic feature.

SUPRASEGMENTAL ORGANIZATION: THE CONTROL OF SPEECH RATE

In the preceding section, variability in the production of a phone due to changes in phonetic environment was discussed. In this section, questions concerning allophonic variation that arises from a different source, the suprasegmental feature of speaking rate, will be explored.

Experiments on the effects of speaking rate and stress have been concerned primarily with the question of whether all such effects can be attributed solely to changes in the timing of commands to the articulators. The classic experiments on the effects of stress and speaking rate on vowels were conducted by Lindblom (1963, 1964). By inferring changes in articulator positions from sound spectrograms, Lindblom found a positive correlation between vowel reduction, or "undershoot," and either decreased stress or increased speaking rate. The failure of the articulators to reach the vowel "target" was attributed to the close temporal succession of motor commands, and so to insufficient time to complete the component gestures. In addition, Lindblom's speaking rate data showed that the rate of target-directed articulator movement remained constant across changes in duration. This supported the concept of undershoot as being a time-based phenomenon; also, it implied a simple model to account for the effects of stress and speaking rate, that is, a cut-off of the commands rather than a complete reorganization of the gesture.

However, in two separate experiments, one cinefluorographic (Gay et al., 1974) and one electromyographic (Gay and Ushijima, 1974), it was found that changes in speaking rate are brought about by a complex reprogramming of the input to the speech string. For example, in a combined electromyographic-cinefluorographic study of speaking rate control (Gay et al., 1974), it was found that an increase in speaking rate was accompanied by a decrease in vowel duration and, for the most part, articulatory undershoot. However, the degree of articulatory undershoot varied with both the individual subject and phonetic identity of the vowel. These differences are illustrated in Figure 7. This figure shows tongue body and jaw movement for the sequence /api/, for two different speaking rates. It is evident that the degree of undershoot is considerably greater for the open vowel /a/ than for the close vowel /i/.

More interesting than the existence of articulatory undershoot for fast speech are the underlying muscle action patterns that control those movements. The EMG data provide a fairly complete account of this control mechanism. These data show that lip muscle activity (orbicularis oris) associated with labial consonant production and tongue tip muscle activity (superior longitudinal) associated with lingual consonant production increase for fast speech, while genioglossus muscle activity for tongue body movement during vowel production decreases during fast speech. These results are illustrated for the sequence /ipi/, in Figure 8.

The first finding implies an increase in articulatory effort and an increase in the speed of articulatory movement: the production of both /p/ and /t/ requires a complete occlusion of the vocal tract, which must be produced more quickly and with greater effort during fast speech. The reduction in EMG activity for the vowel during fast speech, on the other hand, is compatible with the view that a vowel target has a built-in error or tolerance factor that can absorb the extra demands in speech rate.

Two other interesting results appeared in these experiments. One was that lip muscle activity associated with rounding for /u/ also increases with an increase in speaking rate. This implies that the different effects of changes in speaking rate are related either to specific muscle systems or individual phonetic features rather than basic differences in phonetic categories. The other was that, for both subjects, an increase in speaking rate was accompanied by an increase in the frequency levels of both the first and second formants. Thus, since the acoustic triangle is not reduced toward the neutral schwa, articulatory undershoot during fast speech does not produce the same acoustic result as articulatory undershoot during destressed speech.

The most important aspect of the electromyographic speaking rate data is not the direction in which the amplitude of the signals change for consonants and vowels, but rather the fact that they change. Changes in both the timing and amplitude of the EMG signals with changes in speaking rate signified that the control of speech rate requires complex motor programming, and not simply a reordering of the timing function (Lindblom, 1963).

SUMMARY

The major points made in this paper are as follows: first, anticipatory movements toward the second vowel in a VCV sequence begin during the closure period of the intervocalic consonant. This restricted coarticulatory field includes both the tongue body movement and lip rounding gesture associated with the second vowel. Furthermore, the size of this field is not affected by the identity of the intervocalic consonant. Second, like anticipatory effects, carryover effects do not seem to extend beyond an immediately neighbor-

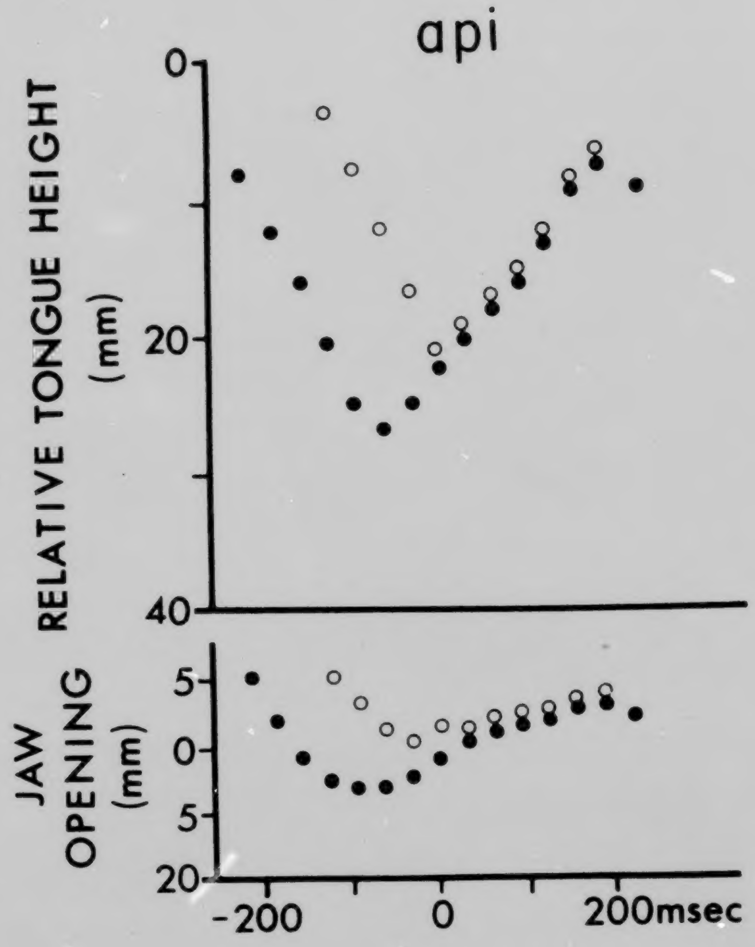


Figure 7: Articulatory movements for two speaking rates. The filled circles correspond to the slow rate and the unfilled circles, the fast rate.

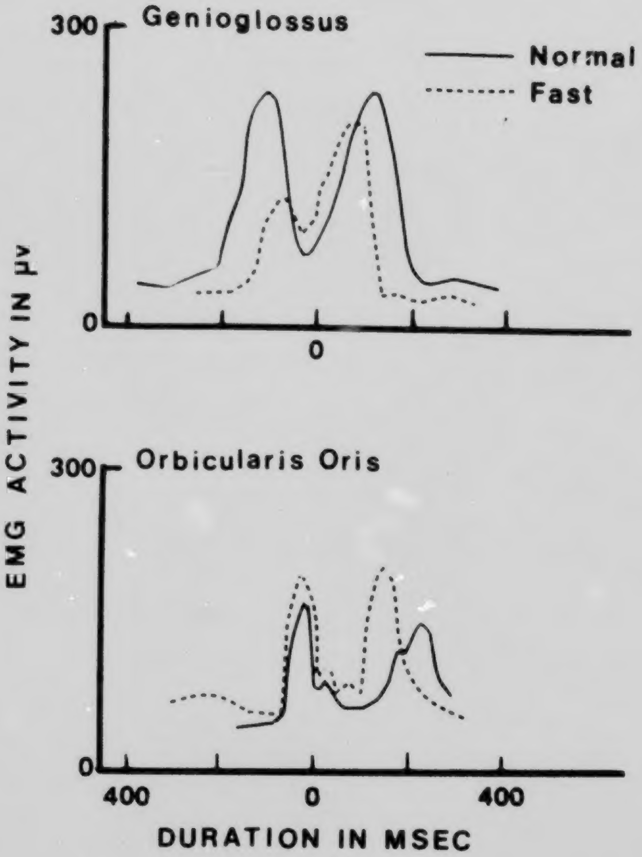


Figure 8: Averaged EMG plots for the genioglossus and orbicularis oris muscles for two different speaking rates.

ing segment. Unlike anticipatory effects, however, the appearance of carryover coarticulation effects depends on the phonetic identity of the particular segment on which these effects might act.

The implication of these findings is that the rules governing the segmental input to a VCV string are probably not as complex as present models suggest. The fact that anticipatory movements begin and primary carryover effects end at about the same time during the closure period of the consonant, suggests that the release of the consonant and movement toward the vowel are organized and produced as an integral articulatory event. This formulation argues against the operation of a scan-ahead mechanism at the segmental level. Rather, all features of both elements of the syllable are contained within the boundaries of that unit.

This does not necessarily mean, however, that a scan-ahead mechanism does not operate at another stage of speech production. The complex reorganization of commands accompanying changes in speaking rate suggests that the temporal features of a downstream segment are known in advance.

Thus, while it has been traditionally considered that the serial ordering of segments is governed by complex rules whose effects can spread across several adjacent segments, and the temporal control of speech is governed by a simple adjustment of timing of commands to the articulators, it may well be that the reverse is true: the segmental input to the speech string is governed by simple rules, while the temporal formulation of the string requires complex articulatory adjustments based on advance information obtained from a higher level scan-ahead mechanism.

REFERENCES

Bell-Berti, F. and K. S. Harris. (1976) EMG study of the coarticulation of lip rounding. J. Acoust. Soc. Am. 60, S63(A).

Daniloff, R. G. and K. L. Moll. (1968) Coarticulation of lip-rounding. J. Speech Hearing Res. 11, 707-721.

Gay, T. (in press) Articulatory movements in VCV sequences. J. Acoust. Soc. Am.

Gay, T. J. (1974a) A cinefluorographic study of vowel production. J. Phonetics 2, 255-266.

Gay, T. J. (1974b) Some electromyographic measures of coarticulation in VCV utterances. Haskins Laboratories Status Report on Speech Research SR-44, 137-145.

Gay, T. and T. Ushijima. (1974) Effect of speaking rate on stop consonant-vowel articulation. A paper presented at the Speech Communication Seminar, Stockholm.

Gay, T., T. Ushijima, H. Hirose and F. S. Cooper. (1974) Effect of speaking rate on labial consonant-vowel articulation. J. Phonetics 2, 46-63.

Henke, W. (1966) Dynamic Articulatory Model of Speech Production Using Computer Simulation. Ph.D. Thesis, M.I.T.

Hughes, O. M. and J. H. Abbs. (1976) Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. Phonetica 33, 199-231.

Kozhevnikov, V. A. and L. A. Chistovich. (1965) Rech', Artikulyatsiya, i Vospriyatiye. Trans. as Speech: Articulation and Perception. 1966.

(Washington, D.C.: Joint Publications Research Service), 30, 543.

Lindblom, B. E. F. (1963) Spectrographic study of vowel reduction. J. Acoust. Soc. Am. 35, 1773-1781.

Lindblom, B. (1964) Articulatory activity in vowels. STL-QPSR 3.

MacNeilage, P. F. (1970) Motor control of serial ordering of speech. Psychol. Rev. 77, 182-195.

MacNeilage, P. F. and J. DeClerk. (1969) On the motor control of coarticulation in CVC monosyllables. J. Acoust. Soc. Am. 45, 1217-1233.

Öhman, S. E. G. (1966) Coarticulation in VCV utterances: Spectrographic measurements. J. Acoust. Soc. Am. 39, 151-168.

Stimulus Dominance and Ear Dominance in the Perception of Dichotic Voicing Contrasts

Bruno H. Repp

ABSTRACT

Two studies were conducted to determine the effect of variations in voice onset time (VOT) on the perception of dichotic stop-consonant-vowel syllables contrasting in the voicing feature. The dichotic stimuli were partially fused, so that only a single response was required. Variations in VOT had a systematic effect on the probability of hearing the fused stimuli as voiced or voiceless. Changing the VOT of a voiceless stimulus had a larger effect than changing the VOT of a voiced stimulus. Unless one of the competing stimuli was close to the category boundary, the perceptual integration of their VOTs seemed to be roughly additive. The relative phase of the periodic portions of the stimuli had an unexpected effect on perception that remains to be explained. A number of subjects showed very strong right-ear dominance in these tests. The range and reliability of the laterality effects obtained, as well as certain other methodological features, make the present tests promising as tools for assessing individual differences in ear dominance.

INTRODUCTION

When two different auditory stimuli are presented simultaneously to the two ears, the perceptual result depends on a number of factors. One of these is dichotic (binaural, lower-level) *fusion*. It determines whether one or two stimuli are heard. If the two inputs are very similar in their spectral and temporal characteristics, they may fuse, so that only a single stimulus is heard. If the two stimuli are dissimilar, two separate events are heard at the two ears, but it may nevertheless be difficult to identify both of them correctly. Two other factors determine which of the two competing inputs is perceptually more prominent. "Perceptual prominence" means, in the case of fused stimuli, that the fused percept sounds more like one component than like the other, or, in the case of unfused stimuli, that one of the inputs is more often correctly identified (or stands out more clearly) than the other. One of the two factors affecting perceptual prominence is *ear dominance*: the

---

Acknowledgment: Experiment I was supported by NIH Grant DE00202 to University of Connecticut Health Center, Farmington, and was made possible by the hospitality of Haskins Laboratories. Experiment II and preparation of this paper were supported by NICHD Grant HD01994 to Haskins Laboratories. I am grateful to Quentin Summerfield for comments on an earlier version and for serving as a subject in Experiment I.

[HASKINS LABORATORIES: Status Report on Speech Research SR-50 1977]

stimulus in one ear may be more prominent than the stimulus in the other ear. The other factor is stimulus dominance: in a particular stimulus combination, one stimulus may be more prominent than the other, regardless of the ear to which it is presented. Finally, it is possible that the perceptual result of dichotic interaction does not resemble either of the two competing stimuli but is a more complex composite of the two--a consequence of higher-level fusion (see Cutting, 1976). Ear dominance, stimulus dominance, and higher-level fusion are probabilistic factors whose effects can be estimated only from a large number of trials; they also vary substantially from listener to listener. Lower-level fusion, on the other hand, is basically deterministic; it may vary in degree between different stimulus combinations, but it usually does not vary over time or between individuals.

The theoretical and methodological significance of these factors for studying dichotic competition between speech sounds has been discussed in several recent papers (Cutting, 1976; Repp, 1976b, 1977a, 1977b). Dichotic listening tasks are used not only to gain insight into the mechanisms of speech perception and selective attention but also--and more frequently--to assess lateral asymmetries in perception that reflect, at least in part, functional lateralization in the brain. In the past, the large majority of studies has used unfused stimuli for these purposes, which were heard as two separate events, so that two responses were required on each trial. Fused (or partially fused) stimuli, on the other hand, require only a single response, which offers certain methodological advantages, especially with respect to measuring laterality effects. There is a close formal analogy between ear dominance and stimulus dominance in the dichotic single-response paradigm on the one hand, and sensitivity and bias in a signal-detection situation on the other hand. As sensitivity can be measured by varying bias and deriving a receiver-operating-characteristic (ROC) function and an associated index of sensitivity (Green and Swets, 1966), so ear dominance can be measured by varying stimulus dominance and deriving an appropriate ROC (isolaterality) function and an associated index (Repp, 1977b). This analogy makes the dichotic single-response paradigm singularly attractive as a tool for measuring laterality effects. The traditional two-response paradigm is much less suited for the application of signal detection procedures.

Several assumptions and prerequisites must be satisfied before the methods of signal detection theory can be effectively applied to the results of a dichotic test using fused stimuli. First of all, ear dominance and stimulus dominance must be independent factors--an assumption that seems reasonable enough to be accepted here without further discussion. Second, any dichotic test contains a number of different stimulus combinations, and it is necessary that ear dominance does not vary as a function of these different combinations, except for random measurement error. In order to satisfy this (testable) assumption, the set of stimuli used in a test should be as homogeneous as possible in terms of their phonetic and auditory properties. Third, each stimulus combination should receive only two different categories of responses, reflecting perceptual dominance of one or the other component stimulus. Stimulus combinations that are subject to higher-level fusion and generate more than two categories of responses are not desirable. Fourth, stimulus dominance in different stimulus combinations should vary systematically over a wide range. Individual subjects often bring very different biases to a task, and a wide variation of intrinsic (that is, average or expected) stimulus dominance relationships makes it less likely that some individuals show such strong biases in most stimulus pairs that their sensitivity (ear dominance) can no longer be measured reliably. In addition, systematic variation of stimulus dominance permits the actual derivation of an ROC function whose shape determines the index of sensitivity to be used. Fifth, for the whole effort to be worthwhile, the stimuli used must generate reliable asymmetries and individual differences in ear dominance. We assume that ear dominance varies in degree between individuals and can be measured on (at least) an ordinal scale (cf. Shankweiler and Studdert-Kennedy, 1975).

The set of stimuli most frequently used in recent dichotic studies is comprised of the consonant-vowel syllables, /ba/, /da/, /ga/, /pa/, /ta/, and /ka/. The fifteen possible combinations of these syllables fall into three sets: place contrasts (/ba/-/da/, /ba/-/ga/, /da/-/ga/, /pa/-/ta/, /pa/-/ka/, and /ta/-/ka/), voicing contrasts (/ba/-/pa/, /da/-/ta/, and /ga/-/ka/), and double-feature contrasts (/ba/-/ta/, /ba/-/ka/, /da/-/pa/, /da/-/ka/, /ga/-/pa/, and /ga/-/ta/). Dichotic (voiced) place contrasts have been investigated in detail by Repp (1976b). When precisely aligned and minimally distinctive synthetic syllables are used, these stimuli fuse perfectly and are virtually indistinguishable from binaural syllables. This makes them ideal for the single-response paradigm. However, they yield only small ear advantages--perfect fusion seems to prevent strong lateral asymmetries. Some of the place contrasts yield a third response category ("psychoacoustic fusion"--cf. Cutting, 1976). Although stimulus dominance relationships can be varied to some degree by changing the acoustic structure of the stimuli, there are limits to this variation, and informal observations have shown that some stimulus combinations have extreme intrinsic biases that are difficult to remove. Thus, place contrasts (voiced place contrasts, at least; voiceless place contrasts have not been investigated in detail) are problematic with regard to three of the requirements named above.

Double-feature contrasts were investigated in detail by Repp (1977a). Precisely aligned synthetic syllables contrasting only in the relevant acoustic parameters (voice onset time, formant transitions) are "partially fused," that is, the perfectly fused vowel portion is preceded by an unfused portion of very brief duration. This unfused portion results from aspiration noise in one ear (the initial portion of the voiceless stimulus) being accompanied by a different noise and/or a periodic waveform in the other ear (the initial portion of the voiced stimulus) and lasts as long as the voice onset time (VOT) of the voiceless stimulus--perhaps 50 msec. The perceptual result is a single stimulus accompanied by some brief noise in one or the other ear. Thus, the single-response paradigm is appropriate for such partially fused stimuli. Repp (1977a) showed that surprisingly large right-ear advantages are obtained in such a test, and that stimulus dominance can be varied by changing the acoustic structure of the stimuli (particularly their VOTs). The only problem with double-feature contrasts is that they lead to a large number of "blend" responses, that is, to four response categories for each stimulus combination (cf. Cutting, 1976). Blend responses are a kind of higher-level fusion and convey no direct information about ear dominance. When calculating ear dominance indices, either a large amount of the data must be discarded, or separate ear dominance indices must be calculated for the voicing and place dimensions, which raises methodological problems, since these indices are often not equal (Repp, 1977a).

Thus, we are left with the voicing contrasts. They are partially fused, like double-feature contrasts, because of the difference in VOT between the two stimuli. For each stimulus pair, there are two different response categories that correspond to the two component stimuli. (For example, /ba/-/pa/ always sounds either more like /ba/ or more like /pa/.) If it could be shown that stimulus dominance can be varied systematically and that large and reliable ear advantages are obtained, voicing contrasts would constitute an optimal dichotic test for measuring ear dominance.

Although the preceding paragraphs emphasized methodological considerations, the present studies were equally motivated on theoretical grounds. (See points 3 and 4 below.) These were the main objectives:

(1) To demonstrate large right-ear advantages and reliable individual differences in ear dominance for voicing contrasts in a single-response paradigm. This was an attempt to replicate the extraordinarily large effects obtained by Repp (1977a) with double-feature contrasts. If it is the voicing distinction (and the resulting partial fusion of the stimuli) that matters, similarly large effects should be obtained with voicing contrasts.

(2) To demonstrate that intrinsic stimulus dominance relationships can be systematically changed by varying the VOTs of the component stimuli. This also amounted to an extension of Repp (1977a), with more attention to detail and to the range over which changes are possible.

(3) To investigate the rule by which competing dichotic VOTs are perceptually integrated. Repp (1977a) obtained a curious interaction: when the voiceless stimulus in one ear had a VOT of +40 msec, a change in the VOT of the voiced stimulus in the other ear from 0 to +15 msec reduced the percentage of voiced responses, as expected; however, when the voiceless stimulus had a VOT of +55 msec, the same change in the voiced stimulus had a slight effect in the opposite direction. The present studies attempted to clarify this interaction by using more steps on the VOT dimension and by factorially combining different VOTs of voiced and voiceless stimuli. Assuming that the interaction would no longer be obtained or could otherwise be accounted for, the question may be asked: According to what rule are competing VOTs integrated into a single percept? Can the process be described by an additive or by a multiplicative model? The theory of functional measurement provides an appropriate framework for this purpose (Anderson, 1974; Massaro and Cohen, 1976).

(4) To investigate the shape of the ROC (isolaterality) function connecting points of equal ear dominance at different levels of stimulus dominance. Since variations in stimulus dominance were to be produced by varying VOT only, and since the place feature was held constant in each test, the stimuli were maximally homogeneous, and constancy of ear dominance with changes in VOT could be safely assumed. Nevertheless, this assumption could be tested by examining the scatter of the data points for individual stimulus pairs. If they do not lie on any single smooth function, lack of homogeneity

would be indicated. If they do, the shape of the function would be of great theoretical and practical interest. Repp (1977b) proposed an index of ear dominance (called e') based on the assumption that the data points (when plotted as "hits" against "false alarms," as described below) follow a linear function that passes through the origin of the unit square, that is, a linear approximation to the standard ROC function of signal detection theory. The e' index was first applied by Repp (1977a), and the data of that experiment supported the assumptions, although there was considerable scatter (see Repp, 1977b, Figure 4). The present studies provided an opportunity for further tests of the assumptions underlying the e' index.

EXPERIMENT I

Method

Subjects. The subjects were eight Yale undergraduates, paid volunteers, some of whom had participated in earlier experiments using synthetic speech and dichotic listening. In addition, the author and a colleague, both highly experienced listeners, participated. Two of the less experienced subjects were left-handers. Three additional subjects were excluded because of poor performance and mistakes that precluded data analysis.

Stimuli. The stimuli were generated on the Haskins Laboratories parallel resonance synthesizer and similar to those used in Repp (1977a). All syllables were 300 msec long, had no initial bursts and a constant fundamental frequency (90 Hz). Different VOTs were generated by setting the amplitude of the first formant to zero and exciting the higher formants with a random noise source for the time specified. There were eight different VOTs: four appropriate for voiced consonants (0, +5, +10, and +15 msec) and four appropriate for voiceless consonants (+40, +45, +50, and +55 msec).

There were two parallel series of stimuli, one containing only labials (/ba/-/pa/), the other only velars (/ga/-/ka/). The stimuli were digitized (with a random sampling error of 0.125 msec) and recorded on tape using the Haskins Laboratories PCM system. Each dichotic series was preceded by 80 binaural syllables--a randomized sequence of the 8 stimuli repeated 10 times. Each dichotic sequence contained 320 stimulus pairs: the 16 VOT combinations (4 voiced stimuli paired with 4 voiceless stimuli) in the two possible channel assignments, in 10 successive individually randomized blocks of 32. The dichotic stimuli were onset-aligned, with a maximal error of 0.125 msec. The interstimulus interval (ISI) was 3 sec.

Procedure. The tapes were played back on an Ampex AG-500 tape recorder, and the subjects listened over Telephonics TDH-39 earphones. The intensities of the two channels were carefully equalized at approximately 75 dB SPL (peak deflections on a voltmeter). The channels were reversed electronically between the labial and velar stimulus series whose order was counterbalanced across subjects. The subjects were not informed about the nature of the dichotic stimuli. Their task was to rate each stimulus heard on a six-point scale. Ratings 1-3 signified /ba/ or /ga/ (1--a clear instance, 3--ambiguous but more like /ba/ or /ga/); ratings 4-6 signified /pa/ or /ka/ (4--ambiguous but more like /pa/ or /ka/, 6--a clear instance). It became obvious in the analysis that the ratings did not provide any significant information beyond that obtained from simply collapsing the ratings into two categories, voiced (1-3) and voiceless (4-6). Therefore, the results are reported here in terms of percentages of voiced responses.

Results

Stimulus Dominance. The stimuli were all reliably categorized in isolation (binaural presentation). The percentages of voiced responses are shown in Table 1. The few errors that occurred reflected the different locations of the category boundaries on the labial and velar continua. The average boundary on a labial continuum, such as the present one, is typically at a VOT of +25 msec, while that on a similar velar continuum occurs around a VOT of +30 msec (Miller, in press). Table 1 shows that the /ba/ closest to the boundary (VOT = +15 msec) received some /pa/ responses, while the corresponding /ga/ was consistently identified. On the other hand, the /pa/s were more consistenty identified than the /ka/s. Because of these differences, dichotic pairs of velar stimuli were expected to receive more voiced responses than pairs of labial stimuli.

TABLE 1: Percentages of voiced responses to the stimuli presented binaurally.

VOT	Labials	Velars
0	100.0	100.0
+5	100.0	100.0
+10	99.0	100.0
+15	93.0	100.0
+40	5.0	5.0
+45	1.0	5.0
+50	0.0	4.0
+55	0.0	6.0

This expectation was borne out: overall, labials received 44.4 percent voiced responses and velars 50.0 percent. However, this effect did not reach significance because two subjects showed a difference in the opposite direction and one subject showed none at all. The overall percentages of voiced responses show that the dichotic stimuli were perceptually well balanced; they were not heard as predominantly voiced or voiceless. Individual differences in the overall percentages of voiced responses ranged from 25.9 to 63.0 percent.

The effects of the variations in VOT are shown in Figure 1, separately for labials and velars. The four functions correspond to the four different VOTs of voiceless stimuli; the VOTs of the voiced stimuli are on the abscissa. Thus, the vertical separation between the functions represents the effect of varying the VOT of the voiceless component, while deviations from horizontali-

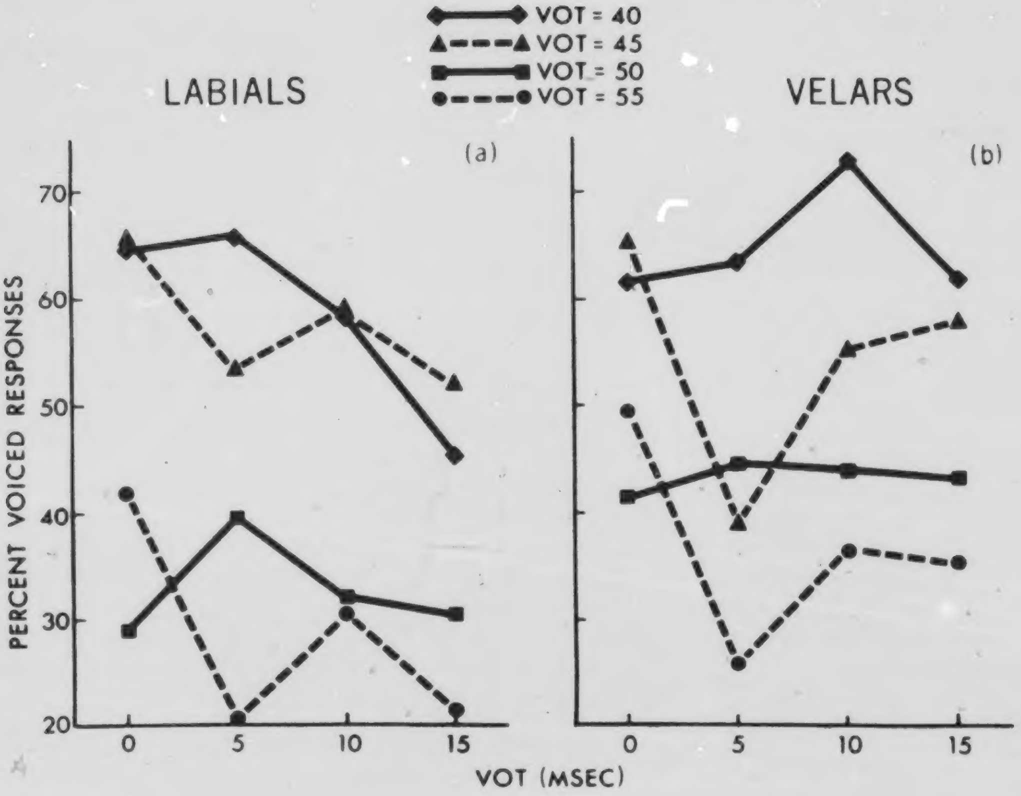


Figure 1: Percentages of voiced responses to the sixteen VOT combinations in the labial (panel a) and velar (panel b) series.

ty represent the effect of varying the VOT of the voiced component. It was expected that the functions would be well separated (with the longest VOT at the bottom), monotonically decreasing from left to right, and parallel (if the two VOT effects are linearly independent and an additive model applies).

The functions obtained were remarkably irregular. Clearly, the VOT of the voiceless stimulus had a pronounced effect ($F_{3,27} = 48.6$, $p << .01$). The effect of the VOT of the voiced stimulus seemed erratic but was nevertheless significant ($F_{3,27} = 6.5$, $p < .01$). The four functions were far from parallel, which was reflected in a highly significant interaction of the two VOT effects ($F_{9,81} = 6.7$, $p << .01$). In addition, each VOT effect interacted with the place (labial vs. velar) factor ($F_{3,27} = 3.0$, $p < .05$, and $F_{3,27} = 7.5$, $p < .01$, respectively).

At first, these results seemed extremely puzzling. The significant effects obtained showed that the irregularities were not just random variation, and it was also obvious that alternate pairs of functions in Figure 1 showed a certain parallellism, although it was not clear why. There must have been some uncontrolled factor in the experiment that influenced stimulus dominance. Although Repp (1977a) also had obtained an interaction between the two VOT effects, the precise pattern of this interaction was not replicated, which added to the confusion.

Eventually, the solution was found in the relative phase of the periodic portions of the dichotic stimuli. The present stimuli had a fundamental frequency of 90 Hz, so that one period of the waveform lasted 11.1 msec. The periodicity began at the VOT specified and continued with a new pulse occurring every 11.1 msec. Since the VOTs were specified in 5-msec steps, the periodic waveforms of some stimulus combinations were nearly in phase, while others were completely out of phase. For example, the fourth and fifth pitch pulses of a stimulus with a VOT of 0 occurred at 44.4 and 55.5 msec, respectively. Thus, this stimulus was nearly in phase with stimuli whose VOTs were +45 or +55 msec, but it was out of phase with stimuli whose VOTs were +40 or +50 msec. Table 2 shows the phase relationships for the different stimulus combinations.

TABLE 2: Phase relationships between the periodic stimulus portions in all dichotic combinations. (Temporal asynchronies between pitch pulses in msec.)

VOT	+40	+45	+50	+55
0	-4.4	0.6	-5.5	-0.5
+5	1.7	-4.4	0.6	-5.5
+10	-3.3	1.7	-4.4	0.6
+15	2.8	-3.3	1.7	-4.4

Note: Negative values indicate that the first pulse of the voiceless stimulus preceded the temporally closest pulse of the voiced stimulus.

It is clear from Table 2 that stimuli whose VOTs differed by a multiple of 10 were out of phase (the maximal pulse asynchrony being 5.5 msec), while the remaining stimulus combinations were more or less in phase. The data of Figure 1 are replotted in Figure 2, separately for in-phase and out-of-phase stimulus pairs. To save space, labials and velars have been combined in this figure.

The transformation of the irregular pattern of Figure 1 into the orderly pattern of Figure 2 is quite remarkable. In particular, it turned out that, once the data were partitioned according to phase, the four functions were nearly parallel within each set of data. In-phase and out-of-phase pairs were analyzed separately in 4-way analyses of variance. Three of the factors were place (labial vs. velar), VOT of voiced stimulus (0 and +5 vs. +10 and +15), and VOT of voiceless stimulus (+40 and +45 vs. +50 and +55). The fourth factor ("VOT shift") represented the effect of simultaneous 5-msec changes in the VOTs of both stimuli--the difference between the solid and the dashed functions in Figure 2.

For in-phase pairs, there was a significant decrease in the percentage of voiced responses as the VOT of the voiced stimulus increased ($F_{1,9} = 11.9$, $p < .01$), and an even larger decrease as the VOT of the voiceless stimulus increased ($F_{1,9} = 36.3$, $p << .01$). The interaction between the two factors was far from significant, which confirms the parallelism of the functions in Figure 2a. The main effect of VOT shift was not significant either. Note that this factor represents here simultaneous 5-msec changes in the VOTs of the two stimuli in opposite directions. Thus, the two VOT effects cancelled each other in this case, despite the fact that, otherwise, VOT changes in the voiceless stimulus had a larger effect than VOT changes in the voiced stimulus. The only other significant effect was a triple interaction between place, VOT shift, and VOT of the voiced stimulus ($F_{1,9} = 8.8$, $p < .05$). It was due to the fact that, in velars, a change in VOT from 0 to +10 resulted in a decrease in voiced responses, but a change from +5 to +15 did not; in labials, the pattern was reversed, if anything (cf. Figure 1).

For out-of-phase pairs, there was a significant effect of the VOT of the voiced stimulus ($F_{1,9} = 17.9$, $p < .01$), but, surprisingly, it went in the opposite direction: the percentage of voiced responses increased with the VOT of the voiced stimulus! The effect of the VOT of the voiceless stimulus was in the expected direction and highly significant ($F_{1,9} = 63.1$, $p << .01$), and so was the effect of VOT shift ($F_{1,9} = 35.9$, $p << .01$). Here, VOT shift represented simultaneous 5-msec changes in the VOTs of the two stimuli in the same direction. The inverted effect of the VOT of the voiced stimulus was apparently not strong enough to cancel the effect of the VOT of the voiceless stimulus. Again, the interaction between the two VOT effects was far from significant, confirming the parallellism of the functions in Figure 2b. There was a highly significant interaction between place and VOT of the voiced stimulus ($F_{1,9} = 22.2$, $p < .01$), and a marginally significant interaction between place and VOT shift ($F_{1,9} = 5.3$, $p < .05$). The first interaction resulted from the fact that the inverted effect of the VOT of the voiced stimulus was entirely due to the velar stimuli (cf. Figure 1); in labial stimuli, the factor had no systematic effect at all. The other interaction was negligible.

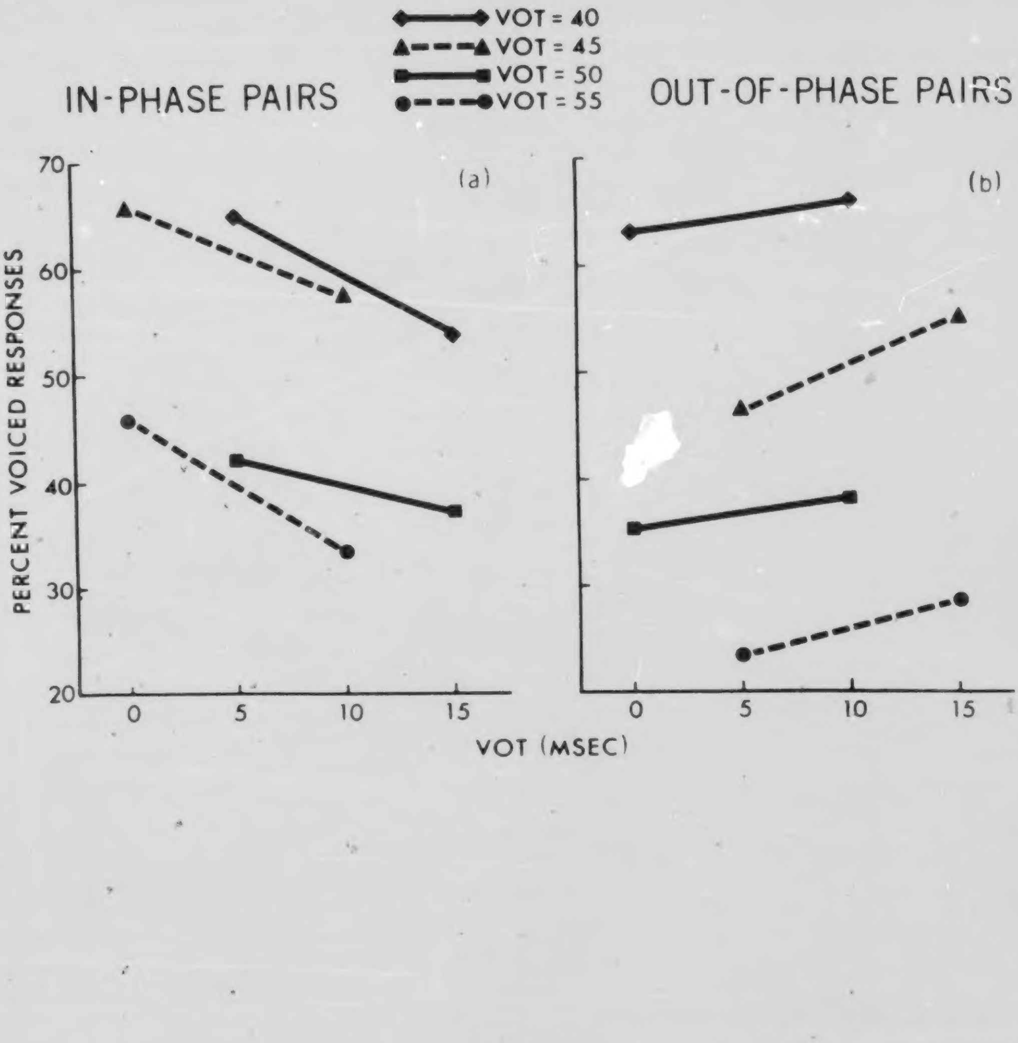


Figure 2: Percentages of voiced responses to in-phase (panel a) and out-of-phase (panel b) stimulus pairs.

Thus, the primary effect of phase was on the effect of the VOT of the voiced stimulus. In addition, in-phase pairs generally received more voiced responses than out-of-phase pairs (cf. Figure 2).

Ear Dominance. The expected large right-ear advantages were obtained. The ear dominance coefficients for the individual subjects are shown in Table 3, separately for the two tests.

TABLE 3: Ear dominance coefficients for individual subjects in the two tests (e' coefficients--see Repp, 1977b).

Subject	Labials	Velars
1[a]	+0.89	+0.87
2	+0.44	+0.08[e]
3	+0.54	+0.96
4[b]	-0.59	+0.01[e]
5[a]	+0.70	+0.71
6	+0.74	+0.75
7	+0.15[e]	+0.62
8	+0.54	+0.59
BHR[c]	+0.95	+0.64
AQS[d]	+0.65	+0.86

[a] Left-handers.

[b] This subject's coefficients may have incorrect signs (see text).

[c] The author.

[d] A colleague.

[e] Not significant. All other coefficients are significant at $p < .05$ or better (see Repp, 1977b, for procedure).

It can be seen that all subjects but one showed large right-ear dominance in at least one test. This includes the two left-handers. (One of them had also participated in Repp, 1977a, and shown a large right-ear advantage there; the same is true for the author). The only large left-ear advantage (subject 4) may represent a mistake in recording the channel-to-ear assignments for this subject. (He was later retested in a dichotic test using similar stimuli and showed a moderate right-ear advantage.) The average ear dominance coefficients were similar for the two tests (+0.50 for labials--+0.62, if the sign of the only left-ear advantage is reversed--and +0.61 for velars) and comparable to that obtained with double-feature contrasts (+0.55--Repp, 1977a).

It is noteworthy, however, that six of the ten subjects showed substantial differences in ear dominance between the two tests. These differences were not related to order of presentation. The correlation between the two tests was +0.70 (but only +0.29 if the sign of the only left-ear advantage was reversed!). This correlation does not reflect low test reliability. Stepped-up test-retest reliabilities (obtained by correlating ear dominance coefficients for the first and second halves of each test and subsequent application of the Spearman-Brown formula--see Lord and Novick, 1968, p.112) for the labial and velar tests were +0.99 and +0.93, respectively (or +0.95 and +0.92, respectively, if the signs of the ear dominance coefficients for subject 4 were reversed). Thus, the correlation between the present tests was distinctly lower than their reliabilities. All correlations were probably somewhat overestimated due to the small number of subjects and the large between-subject variance. A more thorough evaluation of the reliability of the present tests will require a larger subject sample.

The present experiment offered an opportunity to test the assumptions underlying the e' index of ear dominance, as well as the assumption of test homogeneity. The homogeneity assumption says that the data points for individual stimulus combinations--when plotted as "hits" against "false alarms", as described below--should lie on a single, monotonic function, except for random variability. The e' index is based on the assumption that this function--the isolaterality contour or ROC function--is a linear (or slightly curvilinear) function that passes through the origin of the unit square. The unit square is the plot of the proportions of hits against the proportions of false alarms, familiar from signal detection theory.

The ROC function must be symmetric around the negative diagonal, because of the complementarity of the two response categories, each of which may be divided into hits and false alarms. Therefore, it was sufficient to consider only the less frequent response to each stimulus pair, and thus only the area below the negative diagonal of the unit square (Repp, 1977b). For example, if voiceless responses were less frequent than voiced responses for a given stimulus pair, then voiceless responses given that the voiceless stimulus was in the right ear constituted hits, and voiceless responses given that the voiceless stimulus was in the left ear constituted false alarms. These proportions were averaged over all subjects (excluding subject 4) for each individual stimulus combination, so that 32 data points were obtained. They are plotted in Figure 3.

The results were disappointing. The 32 data points clustered in a roughly circular area in the left-hand quadrant of the unit square. The stimuli were homogeneous in so far as all showed sizeable average right-ear advantages. However, variability was so large that it was impossible to determine a single function that fitted the point swarm well. The variation was not systematically related to either the place distinction or the phase factor. All that can be concluded is that there was large, probably random variation between stimulus pairs. For a critical test of the assumptions underlying the e' index, either more observations per stimulus pair or more extreme stimulus dominance relationships are needed.

Discussion

The present study achieved several of its goals. It showed that stimulus dominance relationships in voicing contrasts can be systematically changed by varying the VOTs of the component stimuli, particularly the VOT of the voiceless stimulus in a pair. After taking the phase factor into account, it became clear that the perceptual integration of the VOTs of the two stimuli was approximately linear and additive. Most subjects showed extremely large right-ear advantages, which replicated Repp (1977a). Only the question of

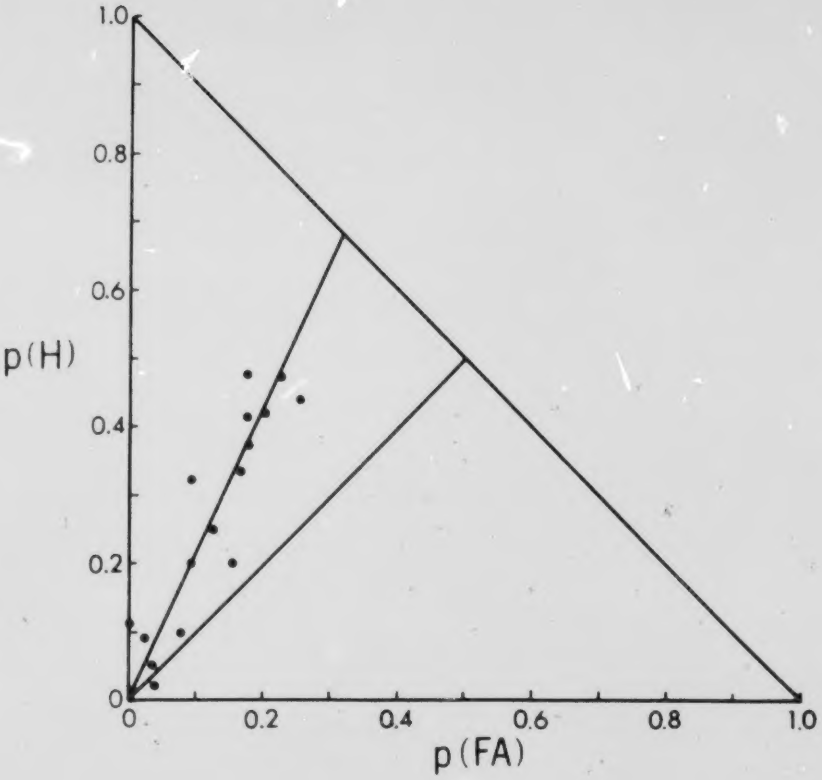


Figure 3: Ear dominance for 32 individual stimulus pairs in Experiment I, plotted as "hits" against "false alarms."

item homogeneity and the shape of the isolaterality contour remained undecided, but at least the results did not directly contradict the assumptions underlying the e' index of ear dominance.

A puzzle was created by the unexpected effect of the relative phase of the stimuli. Why did phase have any effect at all? Why was the effect of the VOT of the voiced stimulus reversed when the stimuli were out of phase? And why did out-of-phase stimuli receive fewer voiced responses? Although phase may be expected to have some effect on fusion, there was no indication that in-phase and out-of-phase stimulus pairs were phenomenologically different. To the author, the test sequences seemed perceptually quite homogeneous, and no stimulus pairs sounded less fused than others.

Note that relative phase applied only to the simultaneous periodic portions of the stimuli, that is, the vocalic portions after voicing onset in the voiceless stimulus. Therefore, it was especially surprising that phase changed the effect of the VOT of the voiced stimulus, since this voicing onset occurred at a time where phase could not yet have played a role. One way of describing the results would be that the voicing feature of in-phase stimuli was determined by a weighted average of the VOTs of the two stimuli; but, in out-of-phase stimuli, it was determined by the weighted difference of the two VOTs. If the difference between the two competing VOTs played a role, an inverted effect of the VOT of the voiced stimulus would be expected, as well as a relative decrease in voiced responses. Perhaps, the decision mechanism responsible for the voicing feature is sensitive to the intervals between any abrupt changes in energy at and shortly after stimulus onset. Normally, there are only two such energy increments: stimulus onset and voicing onset. In the present, partially fused stimuli, however, there were three: stimulus onset, voicing onset in the voiced stimulus, and voicing onset in the voiceless stimulus. Thus, there were three temporal intervals, and the probability of hearing a voiced consonant may have been a weighted function of all three. That intervals other than VOT can affect voicing judgments was demonstrated by Repp (1976a). He found that the interval between stimulus onset in one ear and the onset of an isolated vowel in the other ear had a significant influence on the probability of voiced responses, in addition to VOT. However, there was no indication in these data that the interval between voicing onset in one ear and vowel onset in the other ear played a role, although the relative phase of the dichotic stimuli varied, as in the present experiment. Thus, the phase effects obtained here remain unexplained.

Intriguing as the phase effect was, it was basically an ancillary finding and not essential to the theoretical and methodological purpose of the present experiment. Therefore, it was decided not to investigate the effect further, for the time being, but instead to attempt to replicate Experiment I with stimuli that were definitely in phase. (Even the in-phase stimuli of Experiment I were slightly out of phase.) This was achieved in Experiment II by choosing the VOTs of the voiceless stimuli so that they coincided with pitch pulses in the voiced stimuli.

EXPERIMENT II

Method

Subjects. The subjects were seven paid volunteers and the author. Again, the subjects had had varying degrees of exposure to synthetic speech. One subject was left-handed. The data of one additional subject were rejected because they were too variable.

Stimuli. The stimuli for this experiment were generated on the OVE IIIc synthesizer at Haskins Laboratories, a serial resonance synthesizer that permits finer control of certain stimulus parameters and tends to produce somewhat more "natural" speech. This time, a continuum of alveolar stops was selected, ranging perceptually from /da/ to /ta/. All stimuli were 300 msec long and had a constant fundamental frequency of 125 Hz. This resulted in a period of 8 msec, and the VOTs were spaced accordingly in 8-msec steps. As in Experiment I, there were eight VOTs: 0, +8, +16, +24, +32, +40, +48, and +56 msec. Because of the wider spacing, the fourth VOT (+24 msec) was no longer entirely within the voiced category but fell in the region of the phoneme boundary.

VOT was varied by substituting noise for periodic excitation and setting the first formant to its maximal bandwidth. The pulse generator was turned on at stimulus onset but kept at minimum amplitude during the aspirated portion of the signal; this insured that the first genuine voicing pulse had the intended amplitude. The pulse generator of the synthesizer was synchronized to stimulus onset, so that the first genuine pitch pulse occurred exactly at the VOT specified. Informal observations suggested that an additional factor influencing dichotic stimulus dominance is the relative amplitude of the aspiration noise (which can be controlled in the OVE synthesizer); if it is too low, dichotic voicing contrasts sound predominantly voiced, and if it is too high, they sound predominantly voiceless. In the synthesis specifications, the amplitude setting for the noise generator was selected to be 5 dB higher than the subsequent amplitude setting for the pulse generator. However, since the two amplitude parameters were not on the same scale, the effective amplitude of the aspirated portion was still well below that of the vocalic portion.

The stimuli were digitized at a 10 kHz sampling rate (time-locked to stimulus onset). The stimulus tape contained first a series of 80 binaural syllables--the eight stimuli replicated ten times. A sequence of five blocks of 56 dichotic pairs followed. Each block contained all possible dichotic combinations of the eight stimuli in both channel assignments. In contrast to Experiment I, within-category combinations were included here to facilitate the task by providing unambiguous "anchor" stimuli. The dichotic stimuli were onset-aligned with extreme precision. The ISI interval was 3 sec.

Procedure. The rating scale was no longer used; the subjects simply responded by writing down D or T. After listening to the binaural stimuli, each subject listened to the dichotic series twice. Channels were reversed electronically before the repetition. Stimulus intensity was higher than in Experiment I, about 85 dB SPL (peak deflections on a voltmeter). The more experienced subjects were informed about the dichotic nature of the stimuli.

Inexperienced subjects were simply told to identify the syllables they heard. All subjects were told to ignore any noises they might hear accompanying the stimuli. In the present stimuli, the acoustic segregation of the aspiration noise of the stimulus with the longer VOT was somewhat more noticeable than in Experiment I; if required, it would not have been difficult to tell in which ear this stimulus occurred on a given trial. Nevertheless, it was easy to attend to the fused stimulus in the middle of the head, and the subjects' comments suggested strongly that their responses were not contingent on where they heard the aspiration noise.

Results

Stimulus Dominance. The percentages of voiced responses to the eight binaural stimuli and to the within-category dichotic pairs are shown in Table 4.

TABLE 4: Percentages of voiced responses to binaural stimuli and dichotic within-category combinations.

VOT	0	+8	+16	+24
0	100.0			
+8	100.0	100.0		
+16	100.0	100.0	100.0	
+24	99.4	98.8	95.0	51.3

VOT	+32	+40	+48	+56
+32	1.3	-		
+40	2.5	0.0		
+48	1.9	1.9	0.0	
+56	0.0	1.3	2.5	0.0

It can be seen that all stimuli but one were identified with high consistency. The stimulus with VOT = +24 fell just about at the average category boundary. Most of the errors in voiceless stimulus pairs stemmed from a single subject whose data were somewhat noisy. The most interesting result in Table 4 is that there were hardly any voiceless responses to combinations of the VOT = +24 stimulus with the three voiced stimuli; the boundary stimulus was almost completely dominated by stimuli from within the voiced category.

The results for the between-category combinations are shown in Figure 4. The data were very orderly and confirmed the predictions. The effect of the VOT of the voiced stimulus was highly significant ($F_{3,21} = 47.9$, $p << .01$), and so was the effect of the VOT of the voiceless stimulus ($F_{3,21} = 38.5$, $p << .01$). In addition, there was a significant interaction between the two factors ($F_{9,63} = 11.4$, $p << .01$).

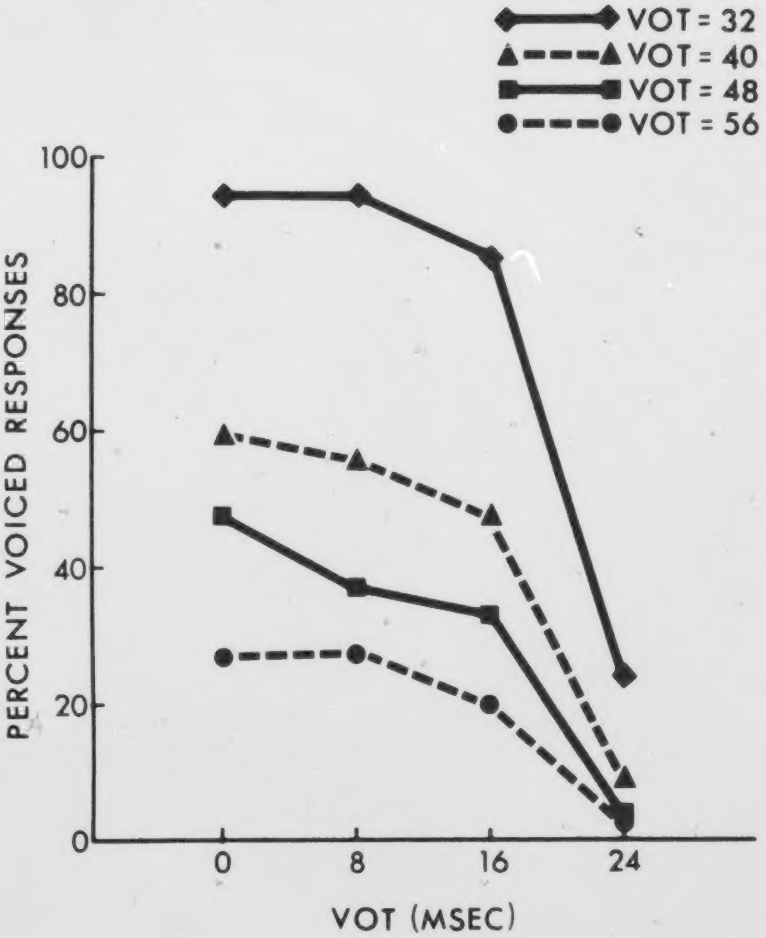


Figure 4: Percentage of voiced responses to sixteen VOT combinations in Experiment II.

The interaction was due to the VOT = +24 stimulus, which was strongly dominated not only by voiced but also by voiceless stimuli. Likewise, the voiceless stimulus closest to the boundary (VOT = +32), although consistently identified in isolation, was strongly dominated by the voiced stimuli and did not completely dominate the VOT = +24 stimulus. Thus, both stimuli flanking the boundary were weak in dichotic competition. The data were reanalyzed omitting these two stimuli. When only combinations of "good instances" of each category were considered, the effect of the VOT of the voiced stimulus was much reduced, but nevertheless in the predicted direction and significant ($F_{2,14} = 7.3$, $p < .01$); the effect of the VOT of the voiceless stimulus was more pronounced ($F_{2,14} = 15.5$, $p < .01$), and there was no longer any significant interaction. In other words, the functions were again parallel (cf. Figure 4).

Ear Dominance. The individual e' coefficients for the eight subjects are shown in Table 5.

TABLE 5: Individual e' coefficients in Experiment II.

Subject	e'
1	+0.14
2	+0.21[c]
3	+0.86
4	-0.34
5	+0.22[c]
6	+0.68
7[a]	0.00[c]
BHR[b]	+0.88

[a] Left-handed.

[b] The author.

[c] Not significant. All other coefficients are significant at $p < .05$ or better.

The average right-ear advantage in this test was smaller than in Experiment I but still substantial (e' = +0.37, based on average scores). Only two subjects and the author showed very large right-ear advantages. Of the remaining subjects, four showed small right-ear advantages and one a moderate left-ear advantage. The reliability of this test was estimated by the split-half method to be +0.96, although some subjects showed considerable variation. The reliability is again somewhat overestimated, due to the small subject sample, but it is nevertheless encouraging.

Figure 5 shows the average hit and false alarm proportions for the sixteen stimulus pairs. It can be seen that, in this test, much more variation in the average "bias" was obtained than in Experiment I--a consequence of including stimuli close to the category boundary. The points are

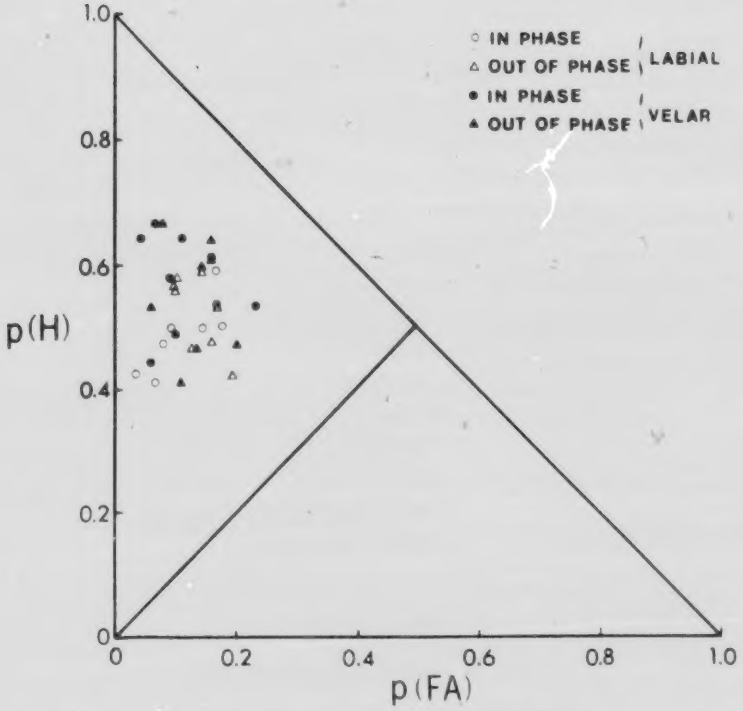


Figure 5: Ear dominance for 16 individual stimulus pairs in Experiment II, plotted as in Figure 3, with the best-fitting linear isolaterality contour drawn in.

clearly fitted best by a linear (or perhaps curvilinear) function through the origin, and not, for example, by a linear function parallel to the positive diagonal. This latter function would be the isolaterality contour corresponding to the simple difference score, p(H) - p(FA), as an index of ear dominance (cf. Repp, 1977b). The present data strongly argue against this simple difference index and support the assumptions underlying the e' index.

Discussion

Experiment II successfully replicated the results obtained in the in-phase condition of Experiment I. The competing VOTs of voicing contrasts appear to be perceptually integrated according to an additive rule, as long as neither VOT is too close to the category boundary. When the VOT of one of the competing stimuli approaches the category boundary, this stimulus loses competitive strength and is dominated by the opponent stimulus. Changes in the VOT of the voiceless stimulus have a more pronounced effect than changes of equal magnitude in the VOT of the voiced stimulus.

The average right-ear advantage in the present experiment was not as large as in Experiment I and in Repp (1977a). However, the exceptionally large average effects obtained earlier were probably fortuitous and due to the small subject samples. These earlier tests probably just happened to include a number of subjects from the upper end of the distribution. Large variation in ear dominance is desirable for methodological purposes: the larger the variation, the more reliable will the measurements be (unless within-subject variability increases in proportion to between-subject variation).

GENERAL DISCUSSION

The effects of variations in acoustic stimulus structure on dichotic stimulus dominance relationships confirm once more that dichotic interaction between speech stimuli does not take place at a purely phonetic level. Effects of variations in VOT on dichotic stimulus dominance have recently also been reported by Carney and Speaks (1976) and Miller (1976). Whether the interaction takes place at a purely auditory level or at an intermediate "multicategorical" stage (Repp, 1976b, 1977a) cannot be decided from the present data. The additivity of VOT effects for true within-category stimuli and the breakdown of additivity in the region of the category boundary seem to be compatible with either possibility. However, the phase effects observed in Experiment I indicate that at least part of the interaction takes place at a strictly auditory level--the same level that determines fusion.

The finding that a change in the VOT of the voiced stimulus had a smaller effect than a change in the VOT of the voiceless stimulus may have been due to the fact that the stimuli increased in amplitude over the first 30 msec or so, which is the region of short VOT values. If this explanation is correct, the effect would constitute additional evidence for interaction at an auditory level. According to informal observations, the intensity of the aspiration noise is another auditory factor affecting stimulus dominance. Strong dominance of voiceless stimuli over voiced ones has sometimes been reported in the literature (for example, Berlin, Lowe-Bell, Cullen, Thompson, and Loovis, 1973). Although these studies used natural speech stimuli, the perceptual asymmetry can almost certainly be explained in terms of the auditory properties of the stimuli (such as long VOTs and heavy aspiration of voiceless stops).

The methodological conclusions from the present studies are very encouraging. The present tests show a pronounced average right-ear advantage, large variation between subjects and high reliability. The data of Experiment II, at least, support the use of the e' index proposed by Repp (1977b). Thus, the present methodology appears very promising for the further investigation of dichotic laterality effects. The size of the average ear advantage (and of several individual ear advantages) obtained here and in Repp (1977a) is without precedent in dichotic research with normal subjects. It may be that dichotic voicing contrasts provide optimal conditions for lateral asymmetries to emerge: they are sufficiently well fused for the single-response paradigm to be used, but not so strongly fused as to suppress laterality effects (cf. Repp, 1976b). Obviously, the small difference between the two competing stimuli in their first 40-60 msec is sufficient to produce strong ear asymmetries. It is intriguing to speculate that there is a direct relationship between dichotic fusion and the suppression of ipsilateral auditory transmission--one of the factors responsible for the ear advantage, according to Kimura's (1961) original theory. The auditory discrepancy at the onset of dichotic voicing contrasts (periodic vs. noise excitation) may lead to very effective ipsilateral suppression. It is interesting in this connection that relative phase (Experiment I) had no effect on the ear advantage, although it affected stimulus dominance. The crucial factor in ear dominance may be the periodic-noise contrast at stimulus onset which, of course, has no particular phase relationship.

It is also likely that dichotic voicing contrasts are not sensitive to selective-attention effects, which constitutes another methodological plus. Although it may be possible to tell in which ear the aspiration noise occurred, the listener would have to infer from this observation that the voiceless stimulus was in the same ear, since there is no clear percept of a separate stimulus. It seems that such inferences can be avoided by proper instructions, unlike the situation with unfused stimuli where two separate events are heard and selective-attention strategies constitute a permanent noise factor that is difficult to control.

One principal question remains: What do the present tests measure, that is, what is their validity? Because of the unusual magnitude of the ear advantages, it is necessary to ask whether the present tests measure the same phenomenon that traditional two-response tests measure. The relatively low correlation between the labial and velar tests in Experiment I is also reason for concern. This finding is reminiscent of the similarly imperfect correlation between the e' coefficients for the voicing and place features in Repp (1977a). Research is now in progress to determine whether different tests and different methodologies assess the same single factor of laterality, or whether perhaps multiple factors are involved. Clearly, there is still much to be learned about measuring ear advantages, and the "perfect test" is still far in the future.

REFERENCES

Anderson, N. H. (1974) Algebraic models in perception. In *Handbook of Perception*, vol. II, ed. by E. C. Carterette and M. P. Friedman. (New York: Academic Press), pp. 215-298.

Berlin, C. I., Lowe-Bell, S. S., Cullen, J. K., Jr., Thompson, C. L., and Loovis, C. F. (1973) Dichotic speech perception: An interpretation of right-ear advantage and temporal offset effects. *J. Acoust. Soc. Am. 53*, 699-709.

Carney, E. and Speaks, C. (1976) Effect of acoustic variation in stimuli on response patterns in dichotic listening. *J. Acoust. Soc. Am. 59* (Supplement No. 1), S6(A).

Cutting, J. E. (1976) Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psych. Rev. 83*, 114-140.

Green, D. M. and Swets, J. A. (1966) *Signal Detection and Psychophysics*. (New York: Wiley).

Kimura, D. (1961) Cerebral dominance and the perception of verbal stimuli. *Canad. J. Psych. 15*, 166-171.

Lord, F. M. and Novick, M. R. (1968) *Statistical Theories of Mental Test Scores.* (Reading, Ma: Addison-Wesley).

Massaro, D. W. and Cohen, M. M. (1976) The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. *J. Acoust. Soc. Am. 60*, 704-717.

Miller, J. L. (1976) Properties of feature detectors for VOT. *J. Acoust. Soc. Am. 59* (Supplement No. 1), S41(A).

Miller, J. L. (in press) The perception of voicing and place of articulation in initial stop consonants: Evidence for the nonindependence of feature processing. *J. Sp. Hear. Res.*

Repp, B. H. (1976a) Dichotic "masking" of voice onset time. *J. Acoust. Soc. Am. 59*, 183-194.

Repp, B. H. (1976b) Identification of dichotic fusions. *J. Acoust. Soc. Am. 60*, 456-469.

Repp, B. H. (1977a) Dichotic competition of speech sounds. *J. Exper. Psych. (Human Percept. Perform.)*, *3*, 37-50.

Repp, B. H. (1977b) Measuring laterality effects in dichotic listening. *Haskins Laboratories Status Report on Speech Research SR-49*, 149-185.

Shankweiler, D. and Studdert-Kennedy, M. (1975) A continuum of lateralization for speech perception? *Brain Lang. 2*, 212-225.

Voicing-Conditioned Durational Differences in Vowels and Consonants in the Speech of Three- and Four-Year Old Children

Lawrence J. Raphael*, Michael F. Dorman† and Donna Geffner††

ABSTRACT

Seven consonant-vowel-consonant minimal pairs, differing only in the voicing characteristic of the final consonants, were elicited from 20 three- and four-year old native speakers of American English. Spectrographic analyses of the utterances revealed that (1) children produced vowel duration differences of the same nature and magnitude as those found in adult speakers' utterances; (2) the duration of a preceding vowel, as well as the duration of voicing during the final consonant closure, are reliable predictors of the voicing characteristic of the final consonant; (3) other measures, such as syllable duration, final consonant closure duration, and vowel duration plus final consonant closure duration, are not as reliable as vowel duration and closure voicing duration as predictors of final consonant voicing; (4) the three- and four-year olds did not produce significantly different vowel or closure voicing duration.

INTRODUCTION

Studies of vowel production by adults have shown that vowels preceding voiced consonants in English are of greater duration than those preceding voiceless consonants (Rositzke, 1939; Heffner, 1941; Belasco, 1953; Peterson and Lehiste, 1960). That is, the vowel of /bid/ is of greater duration than that of /bit/, and that of /bʌz/ greater than that of /bʌs/. These durational differences have been shown to be cues to the voicing characteristic of final consonants in synthetic speech (Denes, 1955; Raphael, 1972). Although it is not yet clear whether these differences in vowel duration have a physiological basis, or are entirely learned behavior, they appear to be robust phenomena in English. Investigators report ratios ranging from 1.25:1 to 2.3:1 between the averaged durations of vowels in opposing voicing contexts (Raphael, 1971).

The purpose of the present study was to investigate the differences in duration between vowels preceding voiced and voiceless stops and fricatives in the speech of three- and four-year-olds. Specifically, answers were sought

---

*Also Herbert H. Lehman College, The City University of New York.

†Also Arizona State University, Tempe, Arizona.

††Also St. Johns University, Queens, New York.

[HASKINS LABORATORIES: Status Report on Speech Research SR-50 1977]

for the following questions:

1. Do young children systematically produce vowels of different durations before word-final voicing oppositions?

2. If such systematic differences occur, are they similar in magnitude to those produced by adult speakers?

3. Is there systematic variation in the duration of voicing during consonant closure?

4. How well do vowel duration and other durational factors predict the intended voicing characteristic of a word-final consonant?

5. Are there age-related differences between three- and four-year olds in the above measures?

METHOD

The twenty informants were nine three-year old and eleven four-year old children. Seventeen of the children were enrolled in a nursery school in the Bronx, New York. The remaining three subjects were children of students and of faculty members of Herbert H. Lehman College of the City University of New York. The age range was from three years and three months to four years and eight months.

Seven minimal pairs were elicited from the children and recorded on audio tape. The seven minimal pairs were: (1) rope-robe, (2) Bert-bird, (3) tight-tied, (4) pick-pig, (5) peck-peg, (6) leaf-leave, and (7) loose-lose. All the English stop contrasts and the two most common fricative contrasts are contained in the list of utterances. The vowel inventory was representative of all categories of tongue height, tongue advancement and tongue "tension."

The experimenters attempted to elicit each word without prompting--that is, without saying the words themselves--as the last element in a syntagmatic response frame or as the response to a picture or object shown, or action performed. Each child was told that he would be playing a word-guessing game and was given six practice items that have not been incorporated into the data. Once the subject seemed to understand the task, the elicitation of the minimal pairs listed above began. With the exception of the second pair on the list (Bert-bird), no minimal pair items were elicited consecutively.

If a child did not respond, or if the child's response was not the target word, then the experimenter spoke the word aloud and said that he would ask again later to see if the child remembered it. The child did fairly often, and the word was recorded. Such responses have been noted as delayed imitations, since the word was not said by the experimenter immediately before the child spoke it. If the child did not subsequently remember the word, then the investigator spoke it and asked the child to repeat it. We have noted such responses as immediate imitations. Often the word was elicited still later in the session, and thus several items were recorded first as immediate imitations and then as delayed imitations. Any "correct" response was immediately followed by a request for repetition.

Of the more than 600 tokens that were initially analyzed, 18 percent were immediate imitations, 21 percent were delayed imitations, and 61 percent were unprompted responses. The necessity for prompting varied greatly from word to word. For example, since all of our subjects watch Sesame Street, Bird (from Big Bird) and Bert (from Ernie and Bert) never had to be prompted. On the other hand, very few of the subjects were acquainted with the verbs leave and peck. No systematic differences have been found among the data derived from the unprompted responses, delayed imitations, or immediate imitations. Nevertheless, all of the immediate imitations have been eliminated from the data except for those which were the only instances of a response type for a given subject.

Since we asked for an immediate repetition of any desired response, most of the data is derived from two or more tokens of each response type per child. Those data derived from only one token of a response type generally resulted from the elimination of other tokens because of extraneous background noise that rendered the spectrograms "unreadable," or from various other causes such as sudden upward leaps in fundamental frequency which virtually transformed spectrograms from wide to narrow band and thus removed the formant information which is essential for segmenting and measuring.

Wideband spectrograms were made of each utterance using a Kay Sonagraph, model 6061B. The tapes were played into the Sonagraph at half-speed in order to lower the frequencies of the children's speech, thus making the presence, initiation, and termination of fundamental frequency more easily discernible. This also facilitated the durational measurements by expanding the time scale of the spectrograms.

The following measures, estimated to the nearest 5 msec, were derived from the spectrograms:

1. The duration of the vowel.

2. The duration of the voicing during the final consonant closure.

3. The duration of the closure for the final consonant.

4. The total duration of the vowel plus the final consonant closure.

5. The syllable duration.

6. The total duration of voicing during the syllable.

7. The total duration of voicing during the vowel and final consonant closure. In the case of the stop-initialed syllables, this duration was the same as the total duration of voicing during the syllable.

We will not dwell at length on the difficulties encountered in segmenting and measuring durations of speech sounds in their spectrographic representations. It is clear that vowel duration differences of the magnitude that we have most often found are easily discernible on spectrograms. Specifying durations, however, is considerably more difficult than simply noting that they are there. This is because of the segmentation problem, and the degree of difficulty depends to a great extent on the utterance in question. A pair of the type tight-tied presents minimal difficulty: the burst releases for the stops are clear indications of both the beginning and the end of the syllable; the onset and offset of formants and of fundamental frequency are generally easily visible and adequately define the limits of the vowel and of the voicing during the consonant closure. On the other hand, a pair of the type leaf-leave presents some serious difficulties, among them being the segmentation of the initial consonant from the vowel, and definition of the limits of the final consonant, especially its termination in low-intensity noise. When uncertainty ran too high, utterances were simply eliminated from the corpus. Thus, 25 percent of the originally recorded and spectrographically displayed tokens were not measured. Otherwise, correspondences were sought between all the tokens of both members of a minimal pair, and some recurring acoustic event was used as a landmark for segmentation. For example, voicing during final consonant closure was measured up to the first break in the regular pulsing of the vocal folds as delineated by the low frequency vertical striations on the wideband spectrogram. This landmark was used even though there were many examples of sporadic voicing after the break and occasionally throughout the final consonant closure. Since closure voicing tends to decrease over time, it is not at all certain whether the vocal pulses are audible, even before the voicing break used as a landmark. If any of the pulsing is audible, however, it seems reasonable to assume that it will occur before the transglottal pressure differential has diminished sufficiently to introduce a hiatus in vocal fold vibration.

RESULTS AND DISCUSSION

The average vowel durations for both members of each minimal pair are shown in Table 1. In 95 percent of the tokens, the vowels are of greater duration when preceding voiced consonants than when preceding voiceless consonants. Of the twenty subjects, fourteen showed no reversals of vowel duration for any of the oppositions. That is, the vowel preceding a voiceless consonant for these subjects was never as long or longer than the one preceding the voiced cognate consonant. Two of the utterance pairs (rope-robe and pick-pig) showed no reversals for any subjects, and only one pair (peck-peg) provided as many as two reversals. Of those six subjects who produced vowels of equal or greater duration before voiceless consonants, only one did so for two minimal pairs (Bert-bird and loose-lose), the others reversing in one pair only.

The durational differences found were similar in magnitude to those produced by adults. The range of adult vowel duration ratios from 1.25:1 to 2.3:1 contains fully 90 percent of the tokens in our data. Almost all those outside this range were greater than 2.3:1.

We now consider the other durational measures and how well they predict the voicing of word-final consonants. Table 1 displays the averaged duration of closure voicing for voiced and voiceless consonants for each minimal pair. For each contrast the mean duration is greater for the voiced consonants. In 90 percent of the contrasting tokens the closure voicing is of greater duration for final voiced consonants. Of 132 contrasts, there are only 13 reversals in duration of closure voicing.

TABLE 1: Averaged durations and ratios of vowels and of voicing during consonant closure for the members of each minimal pair.

Minimal Pair	Average Vowel Duration (msec)	Ratio	Average Closure Voicing Duration (msec)	Ratio
Rope	129.7		50.0	
Robe	209.4	1.6:1	112.5	2.3:1
Bert	227.9		48.7	
Bird	321.5	1.4:1	80.8	1.7:1
Tight	221.8		30.5	
Tied	324.5	1.5:1	75.3	2.5:1
Pick	99.8		20.3	
Pig	196.3	2.0:1	48.2	2.4:1
Peck	131.6		24.5	
Peg	198.0	1.5:1	86.3	3.5:1
Leaf	150.8		48.3	
Leave	247.9	1.6:1	89.7	1.9:1
Loose	136.3		57.5	
Lose	231.8	1.7:1	121.8	2.1:1

It is interesting to note that (in Table 1) the ratios of the closure voicing durations from the voiced to the voiceless contexts are consistently greater than those of the averaged vowel durations for each minimal pair. If we assume perceptual significance for these temporal features, then the effect of the difference in ratios would be to maximize the salience of the cue (closure voicing) with the lesser duration.

Although neither vowel duration nor closure voicing perfectly predicted the intended voicing characteristic of the word-final consonants, the combination of the two measures did. That is, the sum of the durations of vowel and closure voicing for each subject was always greater for the member of a minimal pair ending in a voiced consonant than for the member ending in a voiceless consonant.

Figure 1 displays the magnitude of the averaged differences between the durations of vowels, closure voicing and both measures taken together. The averaged differences for the combined measures for each of the contrasts fall between 124 and 160 msec. It is tempting on the basis of these data to speculate that some small range of overall differences is being aimed at by speakers. The range of overall differences for vowel plus closure voicing is less than 36 msec. However, the way in which the total difference is divided between vowel and closure voicing varies considerably from one utterance pair

FIGURE 1

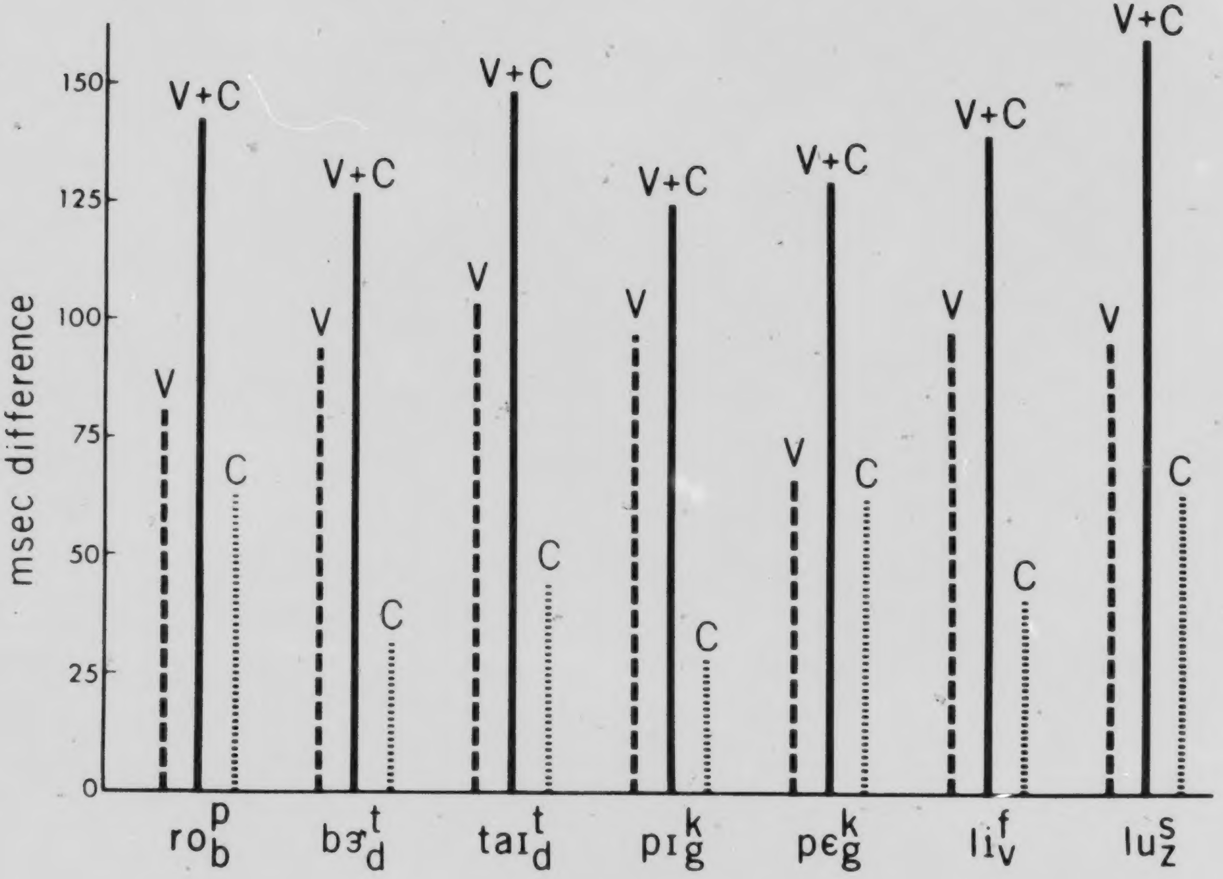


Figure 1: Averaged durational differences between vowels, closure voicing, and the sums of both measures.

to the next. In peck-peg for example, the total difference is comprised by almost equal differences between vowel duration and closure voicing duration. In pick-pig on the other hand, more than three-fourths of the total difference is supplied by the difference between the durations of the vowels.

With one exception, none of the other spectrographic measurements predicted the voicing characteristic of the final consonants as well as either vowel duration or closure voicing duration (or as well as the summed measures). The closure duration of the final consonant was greater for the voiceless cases in 71 percent of the tokens. The sum of vowel duration and final consonant closure duration was greater for 79 percent of the syllables ending in voiced consonants. This last result is to be expected, of course, since the effectiveness of vowel duration as a predictor is reduced by that of closure duration which was most often greater in the case of the voiceless consonants. The total syllable duration was greater for the voiced context in 65 percent of the tokens. The predictive power of this measure suffers because of the variability of the durations of the consonants, especially those in syllable-initial position which varied asystematically with regard to the voicing class of the final consonants.

The exception mentioned above resides in the measure of total voicing duration in the syllable, which predicted final consonant voicing characteristics in 97 percent of the tokens. The strength of this measure exceeds all but that of the perfect predictive power of the summed vowel and closure voicing durations. It falls short of perfection because of the adverse effect of the asystematic variation in the duration of voicing in the initial resonant consonants. In the cases of stop-initiated syllable, this measure was identical with that of the summed durations of vowel and closure voicing.

Finally, we consider age-related differences in the productions of the three- and four-year olds. The average differences between durations of vowels, closure voicing and the two measures combined are shown in Table 2. Although large, the age-related differences are not statistically significant.

TABLE 2: Average differences in msec for vowel duration, final-consonant closure-voicing duration, and both measures combined. Differences are also expressed as percentages of the combined measure.

		Vowel	Closure	Vowel + Closure
Three-year olds	msec	83.2	36.8	120.0
	percent	69.4	30.6	100.0
Four-year olds	msec	98.6	55.6	154.2
	percent	63.9	36.1	100.0
All Subjects	msec	90.2	47.8	138.0
	percent	65.4	34.6	100.0

By age three, children produce differences in vowel and closure voicing durations before voiced and voiceless final consonants which are of a magnitude similar to that found in adults' productions. Research with even younger children will be necessary in order to reveal when such differences are first manifest and how they develop during the first years of language use.

REFERENCES

Belasco, S. (1953) Influence of force of articulation of consonants on vowel duration. J. Acoust. Soc. Am. 25, 1015-1016.

Denes, P. (1955) Effect of duration on the perception of voicing. J. Acoust. Soc. Am. 27, 761-764.

Heffner, R-M. S. (1941) Notes on the length of vowels (IV). Amer. Sp. 16, 204-207.

Peterson, G. E. and I. Lehiste. (1960) Duration of syllable nuclei in English. J. Acoust. Soc. Am. 32, 693-703.

Raphael, L. J. (1971) Vowel duration as a cue to the perceptual separation of cognate sounds in American English. Full volume supplement to Haskins Laboratories Status Report on Speech Research.

Raphael, L. J. (1972) Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. J. Acoust. Soc. Am. 51, 1296-1303.

Rositzke, H. (1939) Vowel length in general American speech. Lg. 15, 99-109.

EMG Signal Processing for Speech Research

Diane Kewley-Port†

ABSTRACT

This paper describes and evaluates the EMG signal processing techniques employed at Haskins Laboratories. EMG signals from many electrodes are collected simultaneously on a multichannel tape recorder along with the audio signal from a speaker. Signals are later processed in two stages. The first includes short-time signal integration. In the second stage, a computer averages a set of EMG signals from repetitions of an utterance carefully aligned to the same acoustic event. The purpose of the signal processing is to produce an average EMG signal that is reliable, relatively noise-free and easy to compare to the corresponding speech events. To evaluate the signal processing, two experiments were conducted in which EMG signals were collected simultaneously from several electrodes placed in the same muscle. Several aspects of the sampling, integration and averaging techniques are evaluated. Results indicate that the signal processing succeeds in producing a reliable, noise-free EMG signal for use in speech research.

INTRODUCTION

Haskins Laboratories has developed an electromyography (EMG) research facility to study speech production. This paper describes and evaluates the signal processing techniques used to transform the small, noisy EMG signal picked up at the electrodes into a reliable, slowly varying signal that can be easily compared to speech events.

Signal processing of the raw EMG signals as designed by Cooper (1965) involves two discrete stages. The first stage is a standard EMG processing procedure of integrating the rectified EMG signal. However, the time constant

---

†Also City University of New York and Indiana University, Bloomington, Ind.

Acknowledgment: I am very grateful to Thomas Baer whose illuminating thoughts and criticisms on earlier drafts are reflected in many places. I thank Katherine S. Harris for careful reading of earlier drafts of this manuscript and for her guidance through the evolution and execution of this work. I am indebted to Dr. Franklin S. Cooper, David Zeichner, Edward R. Wiley and William P. Scully who contributed significantly to the development of the EMG data processing system described here. I also thank Drs. Hajime Hirose and Tatsujiro Ushijima of the University of Tokyo for collaborating in the experiments.

[HASKINS LABORATORIES: Status Report on Speech Research SR-50 1977]

of integration is kept small in relation to the durations of articulatory change. To eliminate the remaining noise, a set of integrated EMG signals from repetitions of the same speech event are averaged together, producing an average EMG signal.

Presumably, the average EMG signal represents the signal for muscle contraction from a small population of motor units. However, in computing the average EMG signal, the randomness inherent in the raw EMG signal is reduced. In this sense, the average EMG signal represents some underlying regularity in the signals causing muscle contraction for a given speech event, which we call the underlying neuromotor command.

This paper is devoted to describing and experimentally evaluating this two-stage EMG signal processing system. These analyses are undertaken to provide at least partial answers to questions commonly raised about this processing technique. Of special interest to the author are the analyses attempting to demonstrate that average EMG signals do represent the underlying neuromotor commands to a muscle.

EMG SIGNAL PROCESSING

Review and Description

The EMG signal is initially detected by electrodes inserted directly in a muscle. Waring (1974), in his review of the use of different electrode technologies, stated that for the purpose of obtaining an estimate of general muscle activity, the type of electrode and the exact specification of its placement in the muscle is not of critical importance. The electrodes in standard use consist of a pair of fine hooked-wires as described by Basmajian and Stecko (1962) and modified for use in speech research by Hirose (1971). Hooked-wire electrodes have been chosen for research of the speech musculature because there is minimal discomfort for the subject with the resulting speech being "normal." Analogous considerations have made similar electrodes an increasingly popular choice in a wide range of applications (Basmajian, 1974).

The electrical signal observed at the electrodes is a bipolar signal that represents the summation of the muscle action potentials from a number of motor units. The motor unit potentials are believed to arise within the volume of fibers within 1 to 2 mm of the electrodes (Buchthal, 1959; Waring, 1974). The single motor unit potential has frequency components up to 2000 Hz (Trimble, Zuber and Trimble, 1973) and a low amplitude, averaging around 500 microvolts, but ranging from 1 to over 1000 μv (Buchthal, 1959; Person and Kudina, 1972; Basmajian, 1974; Waring, 1974). In summation, the potentials from several motor units firing asynchronously partially cancel one another, producing an EMG signal referred to as the interference pattern. The problem then is to recover from this small, jittery, noisy EMG signal, a robust, reliable, slowly moving signal.

The solution adopted here was to pass EMG signals stored on a tape recorder through two stages of signal processing (cf. Cooper, 1965). The first stage is a standard procedure of amplifying, rectifying and integrating the EMG signals. These integrated EMG signals have been shown by various investigators to yield simple functional relationships with measures of muscle tension and velocity of movement under certain experimental conditions. For example, it was found that integrated EMG is linearly related to muscle force (tension) for isometric contractions (Inman, Ralston, Saunders, Feinstein and Wright, 1952; Lippold, 1952), for isotonic contractions (Bigland and Lippold, 1954), for constant torques about the elbow (Leifer, 1969), and in limited unidirectional movement (Bouisset and Goubel, 1971). (Also, see discussion in Basmajian, 1974, p. 168-172). The generality of these findings has been challenged by Zuniga and Simons (1969) for isometric contractions near maximum effort, and by Coggshall and Bekey (1970) for isometric contractions where tension was dynamically developed and tracked. None of these experiments corresponds to the conditions of normal muscular activity in speech where activity is clearly anisometric, and no specific relationship between integrated EMG and articulatory movement of force has been proposed. However, Bell-Berti and Hirose (1975) have shown in one study of velar closure, that the size of the increase in integrated EMG activity and the size of the change in height of the velum were highly correlated, [and Hirose (1976) has shown a similar relationship for PCA activity and the size of the glottal chink].

Integration alone is not a suitable technique to smooth out a noisy interference pattern for comparison to articulatory movement. Significant changes in articulator movement may occur over 5-10 msec for rapidly moving structures (for example, the tongue tip) or 50 msec or longer for slower structures (for example, the jaw). Integration time constants of 5 or 10 msec produce a very noisy looking EMG signal (see Figure 4), and time constants greater than 50 msec severely reduce information concerning signal onset or peak height.

Cooper (1965) proposed a solution to this problem by providing a second stage of signal processing. First, EMG signals are collected from many repetitions of the same utterance and individually integrated with a minimal time constant (5 to 30 msec). Then the digitally sampled EMG signals are computer averaged after careful time alignment of the repetitions.

Model for Averaging EMG Signals

The rationale behind averaging can be discussed using a simple model of the sampled integrated EMG signal. Let the integrated EMG signal after sampling be defined as the sum of two components such that for:

$t = t_k (k=1,2,3,\ldots,n)$, time sampled at instances $t_1, t_2, \ldots, t_n$

$i = i^{th}$ repetition of an utterance

$p = p^{th}$ placement of an electrode in a muscle

$$e_{pi}(t) = S_{pi}(t) + n_{pi}(t) \qquad (1)$$

where

$e_{pi}(t)$ is the waveform sampled by the computer

$S_{pi}(t)$ is the signal representing a continuous underlying motor command

$n_{pi}(t)$ is a random noise signal.

We consider $S_{pi}(t)$ to be the envelope of the integrated EMG signal. The additive noise, then, consists partly of oscillations about this envelope, due to the limited number of motor-units sampled and to the way in which the continuous underlying command is transmitted by discrete motor-unit firings. This component of the noise is defined as zero mean. Another component of $n_{pi}(t)$ is electrical system noise, or the electrode noise sources discussed by Waring (p. 246-48:1974). Due to rectification and integration, this component has a nonzero DC value. However, this component is stationary, so we can redefine the noise as a zero mean by shifting the DC value of the signal. This fact must be remembered when later interpreting the results: only changes in signal level and not absolute levels are meaningful.

In our model, both the ensemble of signal components, $\{S_p(t)\}$ and the ensemble of noise components $\{n_{pi}(t)\}$ are random processes. The main source of randomness in $\{S_p(t)\}$ is normal variation in the production of individual repetition of an utterance, that is, differences in timing and effort of articulatory gestures. As we have defined $\{n_{pi}(t)\}$, it is zero mean over intervals for which the signal is shown moving, although it is not stationary or independent of the signal.

Returning to the model, let $E_p(t)$ stand for the average EMG signal at electrode p, which is calculated as

$$E_p(t) = \frac{1}{m} \sum_{i=1}^{m} e_{pi}(t)$$

for m utterance repetitions and for all t_k samples. Expanding equation (1) for averaging,

$$E_p(t) = \frac{1}{m} \sum_{i=1}^{m} S_{pi}(t) + \frac{1}{m} \sum_{i=1}^{m} n_{pi}(t) \qquad (2)$$

Obviously, for large enough m, the noise term vanishes so that,

$$E_p(t) = \frac{1}{m} \sum_{i=1}^{m} S_{pi}(t) \qquad (3)$$

Success in eliminating the noise depends on a large enough m, which in this case is the number of times a subject repeats an utterance. The filtering effect of integrating $e_{pi}(t)$ using small time constants of 5 to 30 msec reduces the amplitude of the noise and therefore reduces the number of repetitions needed.

Summarizing the two stage signal processing used:

1. EMG signals are collected from many repetitions of an utterance and recorded on a tape recorder.

2. EMG signals are amplified, rectified, integrated with a small time constant, and stored in a computer. This integration, while reducing the high frequency components of the signal, preserves the slow moving components of EMG activity comparable to those of articulatory change.

3. Integrated EMG signals from the same utterance are averaged in time. The average EMG signal, in one model, represents the average of the potentials of a number of motor units firing.

This two-stage process produces a good estimate of the envelope of the interference pattern without excessive time-smearing or too many repetitions.

INSTRUMENTATION

Data Collection and Playback

A block diagram of the recording and playback equipment is presented in Figure 1. The first step in data collection is, of course, the placement of electrodes and, occasionally, pressure transducers. The subject is made as comfortable as possible, and each channel is checked for correct placement and adjusted for maximum gain before the experiment begins. The subject is then asked to read (or repeat) the desired speech material.

The EMG signals collected by the bipolar wire electrodes go to differential preamplifiers that have gains of 40 dB, noise levels (referred to the inputs) of 5 µv RMS, and ca. 100 dB common mode rejection. From the preamplifiers, the signals go to distribution amplifiers with adjustable gains that are usually set at about 30 dB. These amplifiers include 80 Hz high-pass filters with 24 dB roll-off to reject movement artifacts and hum. The filtered signals are then recorded on a one-inch, 14-channel instrumentation recorder (Consolidated Electrodynamics VR-3300). The EMG and pressure signals are recorded in FM mode with an upper frequency limit of 2500 Hz; voice channels and timing and code pulses are recorded in direct mode.

A calibration signal (300 µv ± 1%) referred to the pre-amp inputs is occasionally substituted for each of the physiological signals, so that the signals can be assigned absolute values of microvolts for EMG channels and centimeters of water for pressure channels.

A recording channel is used for voice signals, both for the subject's utterances, and in order to take note of events and changes in procedure during the course of the experiment. Two other channels are used to record a clock track and a code and timing track. The former consists of pulses at a rate of 3200 Hz; the latter, to pulses at a rate of 50 Hz, counted down from the clock. A 4-digit octal timing code is periodically (usually one per second) incremented and superimposed on the 50 Hz pulse train. (This way of introducing the identification codes has provided a good compromise solution to the problem of making the oscillographic record easily readable by humans and the tape-recorded version readable by computer.)

Two data channels can be used for pressure recording. The differential pressure transducers used are SETRA Systems Model 2364. The sensitivity range

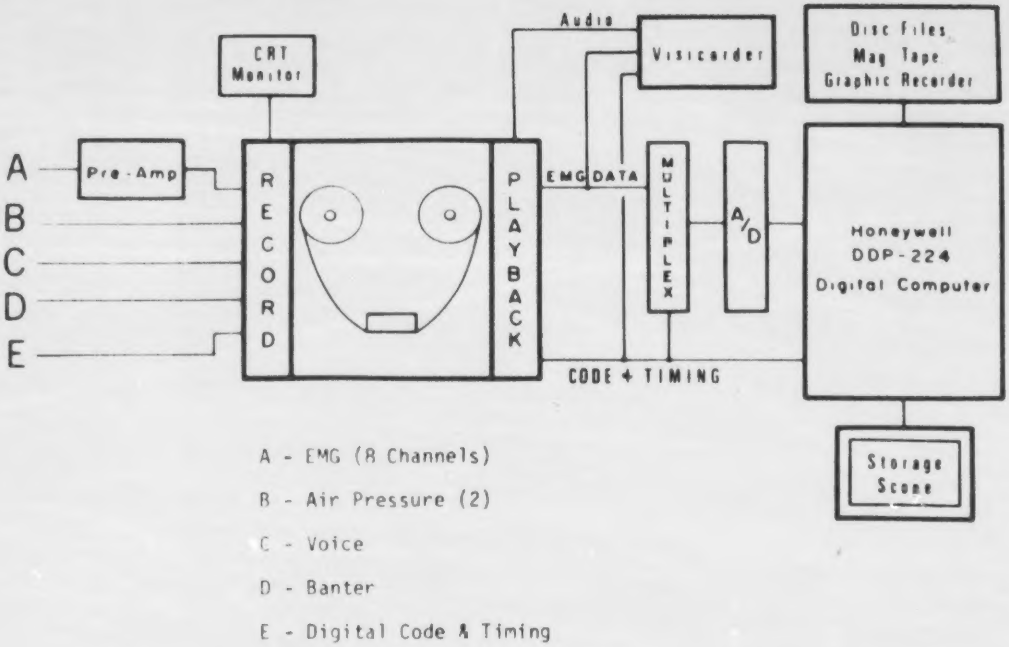

CRT Monitor
Audio
Visicorder
Disc Files
Mag Tape
Graphic Recorder
A
B
C
D
E
Pre-Amp
RECORD
PLAYBACK
EMG DATA
MULTIPLEX
A/D
Honeywell DDP-224 Digital Computer
CODE + TIMING
Storage Scope
A - EMG (8 Channels)
B - Air Pressure (2)
C - Voice
D - Banter
E - Digital Code & Timing

Figure 1: Block diagram of the instrumentation at Haskins Laboratories used in recording and analyzing EMG signals.

extends from 0 to 32 centimeters of water at 4° C, although the usual range is from 0 to 16 cm. water. The frequency response of the system for measuring pharyngeal pressure is from 0 - 300 Hz; the bandwidth of the system for measuring subglottal pressure is somewhat less. Output from the pressure transducers goes directly to dc amplifiers and is recorded on FM channels on the tape recorder. The 300 μv calibration signal is equivalent to 3 cm. water.

For visual inspection of the recorded signals, the EMG data channels, voice channel, and code and timing track are played back as input to an 18-channel Honeywell Visicorder. The recording speed of 7.5 inches per second is doubled on playback to accelerate processing. Each EMG channel is routed again through the distribution amplifiers and high-pass filters, resulting in an overall 80 to 1250 Hz frequency range. The pressure channels are not filtered.

From the amplifiers the signals are full-wave rectified and integrated by linear-reset integrators. These integrators sum linearly over a 5 msec interval, and are reset over 3 msec. They are sampled just before reset by a 16 channel multiplexer. All these operations are controlled by the clock track.

Computer Processing

The basic tasks accomplished by the computer processing system are to locate each utterance on the instrumentation tape, sample and store the EMG signals for each token, and then to align repetitions of the same utterances in time and calculate the average EMG signal.

Analysis of an experiment begins with an inspection of the Visicorder traces. An utterance is located with respect to an octal code recorded on the code and timing track trace. Utterances are usually averaged with respect to a particular acoustic event, such as the beginning of vocal cord vibration, although, of course, utterances could be averaged with respect to another type of event, such as the onset of activity in a particular muscle. A temporal offset interval is measured from the beginning of code to the line-up point using the 20 msec timing pulses on the code and timing trace. Typically, the offset interval is specified to the nearest 5 msec (a quarter of an interval on the timing trace), which is within the inherent uncertainty--estimated at ca. ± 10 msec--that is involved in locating a lineup point on the voice trace. The sampling interval is also 5 msec.

Thus, two descriptors specify an utterance, its CODE and LINE-UP INTERVAL. A sample window of two seconds or less is specified for each utterance relative to the line-up point. A list of the CODES and LINE-UP INTERVALS is prepared for each utterance type to be averaged together by the computer. Two alternative procedures, with equivalent results, are used for making these measurements. Either they can be made by hand, and later typed into a computer, or they can be made directly on a digitizing tablet.

A brief description of the computer configuration and the EMG programs provide an overview of data processing. (Details are available in Kewley-Port, 1973). The computer is a Honeywell DDP-224 with 24 bits/word and a 32K memory. Essential storage devices to permit one-pass sampling of the EMG signals include four 2.4 million word disk units and two high speed magnetic tape units. Programs are supervised by a time-sharing monitor with communication through one of three alphanumeric CRT Terminals. Output devices include a line-printer and a 12 inch storage display scope attached to a hard copy unit.

For purposes of processing, an EMG experiment is defined by the fixed storage capabilities of the computer programs. In a single pass of the EMG analog tape, 8 channels of EMG or pressure data can be simultaneously sampled and stored for up to 30 utterance types, with up to 30 tokens of each type. All the information for one experiment is stored on one digital mag tape. For the purposes of computer processing, the pressure channels are treated like the EMG channels.

Nine computer programs are central to the EMG data processing system. They are listed in order of their use with a brief task description.

E$MGESEL: All control information is input and stored on a mag tape.

E$MGSCAN: EMG signals are sampled as specified by the control information. Information is displayed on the storage scope to enable the operator to optimize playback levels of each channel for the dynamic range of the A/D converters and check for gross clerical errors in the control information.

E$MGSTOR: All EMG signals are sampled with 12 bit precision, stored on disk and then transferred to the mag tape.

E$MGSUMS: The sums and sums of squares of the tokens for each utterance are calculated and stored on map tapes. Only 7 of the 12 bits of data stored are summed. Further digital smoothing of the EMG signals before averaging can be specified in 5 msec increments, called the time constant of integration. The smoothing function is triangular.

E$MGPAGE: All of the individual EMG signals used in calculating an average can be displayed on the storage scope for visual inspection and editing by the experimenter (see Figure 4 for a sample display). The experimenter is usually looking for two types of errors: an EMG signal may appear obviously shifted in time relative to the line-up point for some CODE, in which case the ESEL list can be checked for clerical errors; occasionally very large spikes may be present in the EMG signals that are apparently due to shorting of the electrode tips in the muscle during recording. At the end of the visual editing process, E$MGPAGE automatically constructs a corrected digital mag tape that is then used for all output programs.

E$MGPRNT: The averages and the standard deviations for all utterance types are printed on a line printer from the mag tape. The output is in microvolts using a conversion factor calculated from the 300 microvolt calibration signals that have also been sampled for each channel. For the pressure channels the output values are in cm. water.

E$MGRAPH: The task of rapid and easy retrieval, visual comparison and manipulation of average EMG signals from up to 40 EMG experiments is done by

E$MGRAPH using the storage scope. Up to nine EMG signals may be displayed using three distinctive line types on three sets of axes. The storage scope display can be duplicated on paper by a hard copy unit. (See Figure 5 for a sample display.)

E$MGCORR: A correlation coefficient, r, and its Fisher Z-Transform are calculated and printed for any two EMG averages and/or individual EMG signals on a single EMG experiment. The Fisher Z-Transform is useful in calculating certain statistics on r, especially computing its average (see McNemar, 1969, p. 157-158).

EVALUATION EXPERIMENTS

In order to evaluate the assumptions underlying this data processing technique and to make some other decisions about the details of the analysis procedures, two experiments were conducted. Using a different muscle for each experiment, EMG signals were collected simultaneously from several electrodes placed bilaterally in the muscle. An attempt was made to select muscles that would contract uniformly over their length, using prior experience and anatomical evidence as guides; as we will see below, one of the chosen muscles did not fulfill these conditions.

The subject for both experiments was a female native American English speaker (the author). The basic procedures outlined earlier for collecting and processing EMG signals from hooked-wire electrodes were followed. Any special analyses other than the standard integration and averaging techniques are described below.

In Experiment I, 6 electrodes were inserted bilaterally into the levator "dimple" on the oral surface of the soft palate and one into the orbicularis oris (OO) for reference. After inspection of the processed EMG signal, two recordings were eliminated from the analyses, one because the signal level was very low (below 50 μv), and the other because it contained excessive numbers of erratic, high amplitude spikes thought to arise from the electrodes touching. The remaining channels were two channels from the right levator (LPR2 and LPR3) and two from the left (LPL6 and LPL7). The speech material read 14 times consisted of 4 nonsense utterances, /fimpip/, /fipmip/, /fintip/, /fitnip/ and one anomalous sentence, "Jean Teacup's nap is a snap."

In Experiment II, 6 electrodes were inserted through the cutaneous tissue under the chin bilaterally into the mylohyoid (MY), and one into the orbicularis oris (OO). In this experiment, two MY channels were also eliminated from analysis, both because of very low amplitude signals. The remaining channels were two from the right MY (MYR6 and MYR7) and two from the left (MYL3 and MYL5). The speech material consisted of a text loaded with phonemes for which the mylohyoid was thought to be active, that is, /i/ and apical and velar consonants. The subject read the text, followed by a list of 13 short phrases excerpted from the text, 14 times. Ten 2-second utterances from the text and all the phrases were selected for computer processing.

EVALUATION OF SAMPLING AND INTEGRATION TECHNIQUES

Number of Bits Sampled

Although 12 bits of data are sampled and stored on playback, it was estimated that only 7 bits of data would be reliably sampled because the tape recorder playback amplifiers are limited to approximately 40 dB signal-to-noise ratio. To verify this estimate, a single utterance from Experiment I was sampled and processed with no digital smoothing twice in the same day. Fifty continuous samples were compared for the two passes on all seven channels. The 6 most significant bits were identical so the number of times that the seventh bit differed was counted. This bit differed on the average of 1 in 5 comparisons for all channels. This is equivalent to an average of 6.8 bits reliably sampled, with a range of 6.7 to 6.9 bits across channels. Thus the use of 7 (of the 12) bits sampled in calculating the digital smoothing and averaging of the EMG signals was confirmed as being significant.

Linear Versus RC Integrators

Another purpose of these experiments was to compare the performance of linear-reset integrators with the RC integrators we formerly used in EMG experiments. These RC integrators are, we believe, quite similar to those used by several other investigators. We expected to find a superiority of the linear-reset integrators for several reasons. The RC integrators must be set to a time constant that causes a fixed time lag in the EMG signals relative to other physiological events. For the linear-reset integrators, since the digital integration is both backwards and forwards, there should be no shift of the EMG signals in time. Furthermore, the RC integrators have a filter characteristic that does not provide as good signal integration as the linear integrators, which theoretically provide "true" time integration. The superiority of linear integration for EMG signals in a feedback control application was demonstrated by Kreifeldt (1971) for isotonic contractions in human subjects.

An informal comparison of the output of the two kinds of integrators confirmed the expected superiority of the linear integrators. All EMG signals from both Experiment I and II were processed through both types of integrators. The time constant for digital smoothing was 25 msec in Experiment I, and 35 msec in Experiment II, chosen according to the criteria discussed below. Visual inspection of the averaged EMG curves for the same utterances shows a distinct time lag for the RC integrator signals. Using computer-driven graph pen facilities in E$MGRAPH, the lag was estimated for each of 16 measurements, 11 from Experiment I and 5 from Experiment II. The average lag was calculated as 24 msec in rising portions of the curves, and 21 msec in falling portions. The definition of the EMG curves was sharpened for the linear integrator processing. To estimate the increase in peak height, the difference in microvolts between the peaks of the linear-versus-RC integrator signals was measured for the same 16 curves mentioned above. Calculating the increase as percent of this difference divided by the peak value of the RC curve, the average peak height increase for the linear integrators was 21 percent.

Time Constant of Digital Integration

The time constant for digital integration is not preset, but chosen separately for each experiment. We analyzed some of the data from Experiment II to see if a time constant could be selected, based on the criterion of the minimum time constant necessary for effective smoothing. A number of variables were considered likely to influence the value in different experiments. As mentioned previously, the number of repetitions in the average is an important variable, although the maximum value is set between 15 and 20 due to the limits of endurance of subjects (and experimenters). Other variables to be considered are signal level, and type and length of the speech material. It was apparent from visual inspection of the EMG signals using E$MGPAGE, that more smoothing was needed for signals with higher levels of activity. Thus, two of the mylohyoid electrodes (MYL3 and MYR6) with signal peaks around 500 to 600 µV, were chosen. Two of the 2-sec sentences and two of the 1-sec phrases for the which the mylohyoid showed activity throughout the utterances were selected for analysis. There were 14 repetitions in each average.

The procedure was to calculate the average several times using different time constants. The time constants selected were 5 msec (unsmoothed), 15, 25, 35, 45, 55, and 95 msec. For each time constant, electrode channel and utterance, a correlation analysis was made. The correlation coefficient, r, and its Fisher z-transform, z, were calculated between an EMG average and an individual signal from the average. As the time constant increases, r increases because both signals have less ripple, that is, less uncorrelated noise. To obtain a function representing this increase in r, ten individual EMG signals were correlated with their mutual average for each time constant, and an average of these correlation coefficients was computed using the z values. The increase in r represents the extent to which the individual EMG signals making up an average are becoming more like the average as the time constant is increased.

Figure 2 shows the average r's as a function of the time constant for electrode MYL3; on MYR6 the correlation functions for the four utterances did not overlap as extensively as those of MYL3, but otherwise there was little difference and they are not presented. Average r increased most rapidly for time constants less than 25 msec. The highest value of average r occurred for different sentences for the two electrodes, but a difference in average r functions between sentences and phrases was not observed.

To determine the effectiveness of increased digital smoothing, a function showing the average improvement of r with increased smoothing was calculated. Using the z values, the average z's were calculated for 10 msec increments in time constant. Successive 10 msec pairs were subtracted. The z differences obtained for each pair did not differ for sentences versus phrases, or across channels MYL3 and MYR6. Thus a grand average of z differences was calculated, converted back to r values, and plotted on Figure 3.

The curve on Figure 3 appears to have an elbow about 35 msec, such that some improvement occurs with time constants up to 35 msec, but little improvement for greater time constants. Thus for these EMG signals, a time constant of no more than 35 msec should be chosen; larger time constants of digital integration will probably not remove substantially more ripple from

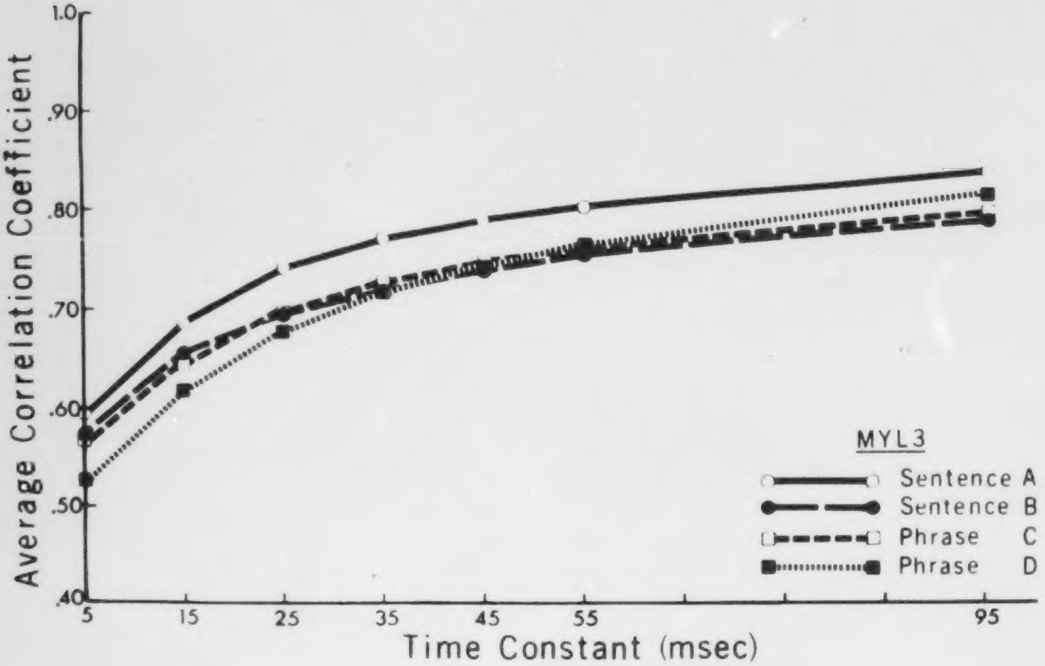


Figure 2: Average values of the correlation coefficients computed between ten individual utterances and their mutual average as a function of increasing the time constant of digital smoothing.

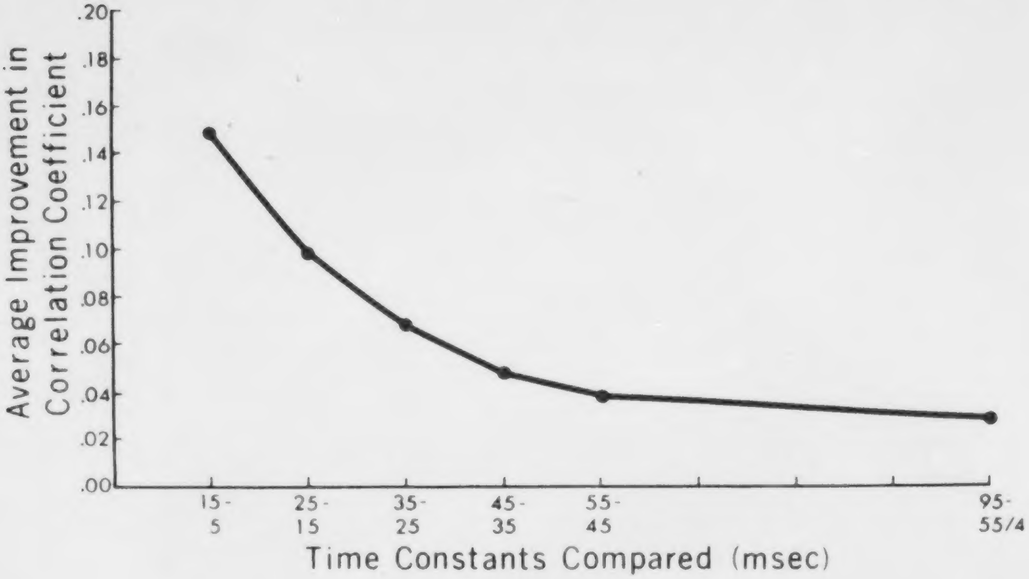


Figure 3: Each point was calculated as the difference between the average correlation coefficients of the two time constants of digital smoothing indicated on the abscissa. See text for more detail.

the average signal, but will certainly reduce the time resolution of the peaks and valleys of the signal.

EVALUATION OF AVERAGING TECHNIQUE

Comparison of Average EMG Signals from Different Electrode Placements

The purpose of calculating an averaged EMG signal for an utterance is to obtain a reliable and relatively noise free EMG signal that we believe is representative of the underlying neuromotor command to the muscle. Using the model previously described (beginning p. 3), we present an analysis that verifies the above statement when EMG signals are observed simultaneously from several electrodes in the same muscle.

Writing equation (3) again, the average EMG signal for a specific electrode is:

$$E_p(t) = \frac{1}{m} \sum_{i=1}^{m} S_{pi}(t) \tag{3}$$

Let's consider the case where a muscle contracts uniformly over its length. In this case, the motor units should fire similarly throughout the muscle (within the normal stochastic limits), and the mean value function represented by $E_p(t)$ should be the same for any electrode placement p. Thus we can define S(t) such that:

$$S(t) = E_p(t) = \frac{1}{m} \sum_{i=1}^{m} S_{pi}(t) \tag{4}$$

for all electrode placements p in a uniformly acting muscle. We can think of S(t) as an EMG representation of the underlying neuromotor command for this utterance.

It might appear that the most obvious way to test whether $E_a(t)-E_b(t) = S(t)$ for electrodes a and b would be to subtract $E_a(t)-E_b(t)$ and see if the resultant function is zero valued. However, amplitude values for any electrode depend in large part on the number and proximity of the motor units firing to the electrode. Rather than try to develop an appropriate normalization technique for signal amplitude, it was decided to use the correlation coefficient, r, which is sensitive to signal variation but not to absolute amplitude, to test the hypothesis. Two analyses are presented, one for correlations between separate EMG electrode averages, and the other for correlations between two groups of repetitions from the same electrode.

To see whether $E_p(t)$ is independent of electrode placement, the correlation coefficient, r_{ab}, is calculated between the averages from two different electrodes, a and b. The correlation coefficient for EMG averages $E_a(t)$ and $E_b(t)$ is defined as:

$$r_{ab} = \frac{\mathrm{Cov}[E_a(t),\ E_b(t)]}{\sqrt{V[E_a(t)]\ V[E_b(n)]}} \quad (5)$$

where Cov stands for covariance and V for variance. But if our hypothesis that $E_a(t) = E_b(t) = S(t)$ is correct, then

$$r_{ab} = \frac{(\mathrm{Cov}[S(t),\ S(t)]}{V[S(t)]} = 1$$

Results from Experiment I

Correlation coefficients were calculated for average EMG signals between all possible pairs of the four electrodes placed in the levator palatini in Experiment I. Table 1 presents the coefficients for all 5 utterances. The coefficients have been grouped according to electrodes that were in the same side of the levator, 2 & 3 and 6 & 7, or different sides, 2 & 6, 2 & 7, 3 & 6, 3 & 7. The average value of the coeffiecients was .95.

TABLE 1: Experiment 1: Correlation Coefficients for Average EMG Signals for Levator Palatini

Electrode Placement		/fimpip/	/fipmip/	/fintip/	/fitnip/	Sentence
Same Side						
	2,3	.93	.96	.96	.95	.95
	6,7	.99	.99	.99	.99	.98
Different Side						
	2,6	.86	.94	.93	.94	.96
	2,7	.86	.93	.94	.92	.93
	3,6	.94	.96	.95	.94	.94
	3,7	.95	.97	.95	.95	.93

The square of the average correlation coefficient, $r^2 = .90$, can be interpreted to mean that 90 percent of the variance of $E_p(t)$ from one electrode is predictable from the variance from another electrode. (Note that variance here refers to the amplitude distribution of one signal about its mean value calculated over the time period sampled.) That is, the hypothesized S(t) signal could account for 90 percent of the variance of the EMG averages over time and other, unknown sources account for the remaining 10 percent of the variance.

One source of this variance might be that the levator is functionally differentiated and there is some real variation in the neuromotor commands to the muscle as observed by the different electrodes. In this regard, note that the two electrode pairs on the same side of the levator had a higher average correlation, .98, than did the four electrode pairs in the different sides, .94. These values are significantly different by a t-test, $t = 4.26$, $p < .001$. This suggests that there was some real variation in the average EMG signals between electrodes, since the results are compatible with the expectation that there would be more variation between the bilateral muscle pairs than within the same muscle.

Results from Experiment II

The results from the mylohyoid muscle observed in Experiment II did not simply replicate Experiment I. It was obvious from visual inspection of the EMG signals that there was substantial variation between the four electrode placements. There appeared to be a functional differentiation of EMG activity for the electrode placements going from the chin posteriorly for speech segments including high front vowels and apical consonants, but uniform EMG activity for the velar consonants. Although this unanticipated finding should be confirmed, it is anatomically reasonable, since the fibers are distributed in their origin and in section-spatially. Correlation coefficients were calculated for the four pairs of mylohyoid electrodes and the results can be interpreted as supporting the above hypothesis. The average of the correlation coefficients for four sentences for adjacent anterior-posterior electrode placements was .92. The average for the extreme anterior-posterior pair for the same sentences was .72.

Further Analysis

The above results from Experiments I and II raise the question: Can a high correlation between two separate electrode signals determine whether or not both signals represent the same underlying neuromotor command? For example, is the average r across all 4 levator electrodes of .95 high enough when there is still 10 percent unaccounted variance between electrodes?

A separate analysis was developed to substantiate when two average signals from electrodes a and b were significantly different from one another. The individual EMG signals, $e_{pi}(t)$, are arbitrarily assigned to two groups of equal size $(\frac{m}{2})$ designated "o" for odd and "e" for even. New averages for each group are calculated, $E_{ao}(t)$, $E_{ae}(t)$, $E_{bo}(t)$ and $E_{be}(t)$. Correlation coefficients are calculated between all pairs of the averages. The correlation coefficients for the odd and even groups from the same electrode are compared by means of a t-test to the correlation coefficients for groups from the different electrodes. If the correlation coefficients do not differ significantly from each other, we may say that the signal variance from one group for electrode a may be predicted equally from the other group on electrode a, or from either group on electrode b. Thus we infer that a single average neuromotor command, $S(t)$, was sampled at both electrode placements.

The EMG signals for the anomalous sentence from Experiment I were analyzed for all four electrode placements in the levator palatini, labeled here as 2, 3, 6, and 7. There were 10 repetitions of the sentence in the EMG averages, each 2 sec long. Figure 4 presents the EMG averages and 4 of the 10 utterance repetitions for electrode 6 (LPL6).

The 10 utterance repetitions were split into odd and even lists on each channel and new averages were calculated. The correlation coefficients for the averages from the pairs of odd and even channels were calculated and appear in Table 2. T-tests were calculated between the means of the correlation coefficients (using the z-Transform) for EMG average from the same electrode, and the averages from the different electrodes. (For example, a t-test was calculated between the mean from the same electrodes, 1/2[r(2e, 2o) + r(3e, 3o)], and mean from different electrodes, 1/2 [r(2e, 3o] + r(2o, 3e)], where e = even and o = odd.) None of the t-tests was significantly different at better than a .25 level (two-tailed test).

TABLE 2: Experiment I: Correlation Coefficients Between the Averages for the Odd and Even Lists for Pairs of Electrodes, 2, 3, 6, and 7.

Even Odd	2	3		6	7		3	7
2	.92	.91	6	.92	.89	3	.89	.79
3	.86	.89	7	.92	.91	7	.92	.91

Even Odd	3	6		2	7		2	6
3	.89	.82	2	.92	.84	2	.92	.88
6	.92	.92	7	.89	.91	6	.89	.91

The EMG signals from one sentence from Experiment II were analyzed in the same way as the sentence from Experiment I. The "sentence" was the first 2 sec from the sentence beginning "Eve and Clayton left Kansas for the ..." The line-up point was the onset of voicing in "Eve." There were 14 repetitions of this sentence. The four electrode placements in the mylohyoid are labeled 3, 5, 6 and 7. As mentioned in discussion of the averaging technique, correlations of EMG averages between pairs of electrodes in Experiment II were not in general high. This sentence was chosen because the coefficients, as seen in Table 3, were higher than most others, probably because there were more velar consonants. However, pairs 3-6 and 3-7 had such low coefficients, .77 and .70 respectively, that they were dropped from further analysis.

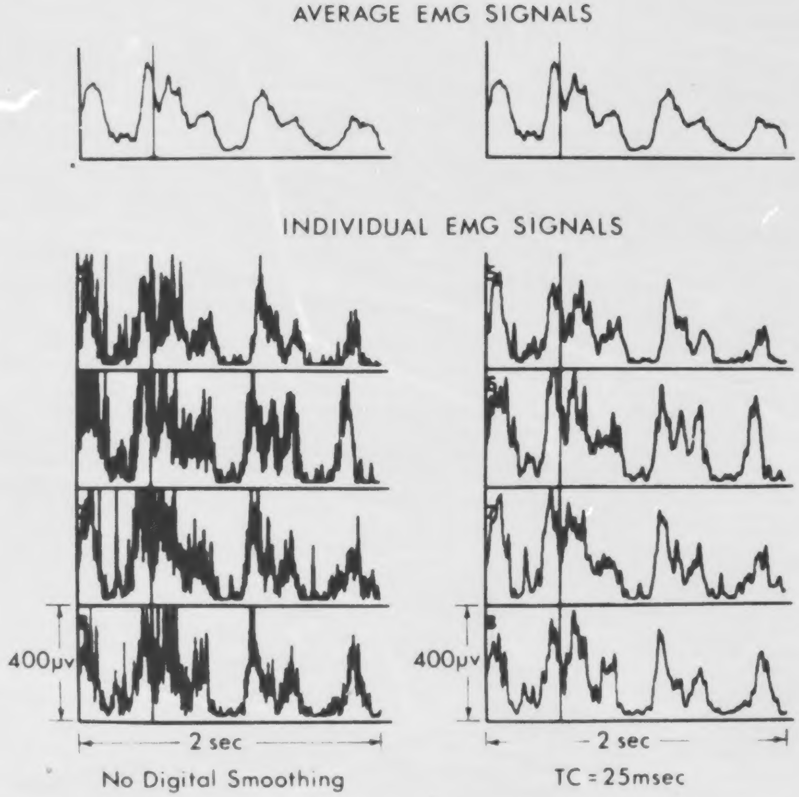


Figure 4: EMG signals from LPL6 for the sentence "Jean Teacup's nap is a snap." The average signal, repeated at the top, is calculated from 10 repetitions using a 25 msec time constant. The line-up point is the offset of voicing in Jean. On the left, sentence repetitions 5 through 8 have no digital smoothing. On the right, repetitions 5 through 8 have a 25 msec time constant.

TABLE 3: Experiment II: Correlation Coefficients Between the Average EMG Signals for Pairs of Electrodes 3, 5, 6, and 7.

Electrodes	5	6	7
3	.87	.77	.70
5		.92	.89
6			.97

The utterances were divided into odd and even lists and the appropriate EMG averages calculated for the remaining electrode pairs. Correlation coefficients between these averages are presented in Table 4. Calculations of t-tests between electrode pairs showed that two pairs of EMG averages were significantly different from one another; t(3-5) = 8.49, $p < .02$, and t(5-7) = 5.70, $p < .05$. For the other two pairs, t-tests showed them to be significantly different at the $.1 < p < .05$ level; t(6-7) = 3.30 and t(5-6) = 3.56. However, both of these calculations are based on correlation coefficients with four significant digits because the rounded two digit entries in Table 4 were identical. It seems that the t-test is not valid in this case since the population variances of the same and different pairs are most likely unequal (see McNemar, 1969, p. 117). Intuitively, it seems that pair 6-7 with an $r = .97$ between the average signals should not be significantly different using this odd-even analysis, contrary to the above which showed a difference at the .1 level. Therefore, if the arbitrary division of EMG signals (here as odd or even) results in identical r's for either the same or different sides, the t-test should not be used, and perhaps a different division of the lists would result in unequal r's. Unfortunately an analysis of pairs 6-7 and 5-6 was not carried out for a different grouping.

TABLE 4: Experiment II: Correlation Coefficients Between the Averages for the Odd and Even Lists for Pairs of Electrodes 3, 5, 6, and 7.

Even Odd				
	3	5	6	7
	.92	.83	.95	.92
	.80	.93	.93	.95

Even Odd				
	5	6	5	7
	.93	.88	.93	.85
	.88	.95	.84	.95

Discussion

Thus, the results obtained from Experiment II showed that for electrode pairs 3-5 whose average signals correlated with an $r = .87$ and pair 5-7 whose $r = .89$, the EMG averages can be considered significantly different from one another. In Experiment I, the r's for the averages from all 6 pairs of electrodes ranged from .93 to .98, and all pairs were shown not be significantly different in the odd-even analysis. A thorough study of many electrode pairs, several utterances and different muscles using the odd-even analysis could possibly determine for which high values of r two EMG signals can be considered the same.

Sources of Variation Removed by Averaging: Speaking Rate Variation

Although the averaging of EMG signals produces a smooth signal, this technique removes not only the noise component but also variation in the individual EMG signals that might be of interest. We know that different repetitions of an utterance are not identical. For purposes of discussion, consider that there are two sources of variation eliminated by averaging, that arising from normal utterance variation and that arising from the noise component, $\{n_p(t)\}$. The purpose of the following studies was to examine the effects of normal speaking rate variation in the averaging technique.

EMG signal averages are calculated after all the repetitions of an utterance are lined up with reference to a single acoustic event. When speaking rate variation occurs, signal peaks will become more unsynchronized the further away they are from the chosen acoustic event. A small experiment was conducted to examine the effects of speaking rate variation on the peaks of the averaged EMG signal.

The anomalous sentence from Experiment I, "Jean Teacup's nap is a snap," was separately processed through all of the programs using three different line-up points: the first line-up point was the onset of voicing in "Jean," the second was the onset of voicing in "snap," and the third was the end of voicing after "Jean." For all four levator palatini electrodes, the effects of changing the line-up point were the same (see Figure 5 for signals from one electrode). Peaks became lower and broader as they were further removed in time from the line-up point. For example, seven peaks are clearly observable in Figure 5A and 5C, but three of them merge into one in Figure 5B where the line-up point is at the end of the sentence. The decrease in peak height was not uniform when the line-up point was moved from one end to the other. The average decrease was calculated as percent of (higher peak - lower peak)/(lower peak) for two channels, LPR2 and LPL6. For the line-up point extremes seen in Figures 5A and 5B, the average percent decrease was 13 percent for the first peak and 28 percent for the last peak. Of course, the extent to which a peak is flattened will depend on its breadth in the original utterances. A brief event will be more affected than one that lasts longer.

The following qualitative generalizations may be drawn from these results. The determination of the onset or offset of EMG activity cannot be accurately made from the EMG average signals when the distance from the line-up point is greater than about one second. A sampling window for averaging of greater than 2 seconds is not advisable.

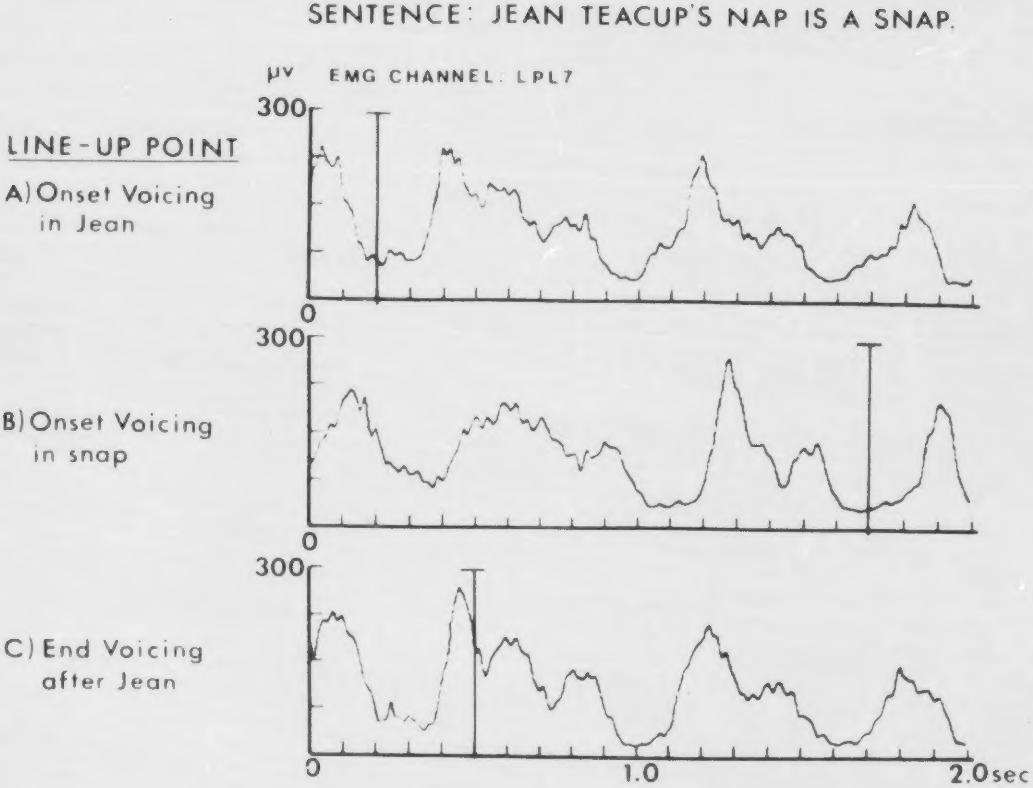


Figure 5: Average EMG signals for one sentence analyzed three times, each with a different line-up point.

Since effects of speaking rate variation are observed in the average EMG signal, we would like to know if the magnitude of these effects is so large that timing information in the average EMG signal is seriously compromised. This question can be partially answered by considering how closely the time course of the individual EMG signals resembles that of the averaged EMG signal. (Timing variations between individual EMG signals can be observed in Figure 4.) Again the correlation coefficient, r, may be used to quantify this resemblance because r is not sensitive to absolute amplitude.

The same procedure described in section "Time Constant of Digital Integration," was used. For the utterances and electrodes shown in Table 5, the correlation coefficients for each individual EMG signal are calculated with their mutual average EMG signal. From Experiment I, the correlation coefficients were calculated for 10 tokens of /fimpip/ and 10 of the anomalous sentence for two electrodes, 2 and 6. From Experiment II, results were taken from Figure 2 for the time constant value 35 msec that was standard for this experiment. The average correlation coefficient and the corresponding range of correlation coefficients for each of 10 tokens were calculated for all 4 utterances appearing in Figure 2.

TABLE 5: Average Correlation Coefficients for EMG Signals from Single Utterances Compared to Their Mutual Average EMG Signals and Their Corresponding Ranges.

Experiment I, TC = 25 msec

	/fimpip/	/fimpip/	Sent.	Sent.
Electrode	LPR2	LPL6	LPR2	LPL6
Average r	.85	.88	.81	.84
Range r	.82 - .90	.84 - .91	.71 - .86	.77 - .99

Experiment II, TC = 35 msec

	Sent. A	Sent. B	Phrase C	Phrase D
Electrode	MYL3	MYL3	MYL3	MYL3
Average r	.77	.72	.72	.72
Range r	.66 - .85	.62 - .84	.58 - .84	.58 - .78

The closeness of the resemblance of a single EMG signal with its average can be ascertained from Table 5. The correlations are somewhat higher for the nonsense words than the sentences, presumably because the sentences are longer. The square of the correlation coefficient, r^2, indicates the proportion of the variance in the average EMG signal that can be predicted from the individual EMG signals. The remaining variance, $1-r^2$, arises from differences in the time course of the two signals, due to speaking rate variations and the remains of the smoothed, uncorrelated noise component $n_{pi}(t)$. In this analysis, we cannot tell what percent of the uncorrelated variance arises from the speaking rate variation, and what percent from the noise component. Calculating r^2 from Table 5, for Experiment I, the proportion of variance predicted ranges from .77 to .66, and for Experiment II, from .59 to .50. Thus the signal variance is predictable from the individual EMG signals.

SUMMARY

This report describes a two-stage EMG signal processing system developed for speech research. In the first stage, EMG signals from a number of repetitions of the same utterance are amplified, rectified and integrated. In the second stage, the integrated signals are carefully aligned in time and averaged, producing an average EMG signal. A model of the integrated EMG signal is described as the sum of a signal representing the motor unit potentials specifying muscle contraction and a random noise signal. This simple model is used in several correlation analyses of the integrated and averaged EMG signals to determine the reliability of average EMG signals and the sources of variance removed by averaging.

Two experiments involving multiple electrode placements in the same muscle provided data for the analyses. Some of the conclusions discussed may be stated as follows:

Because of their precision, linear-reset integrators having a small time-constant of 5 msec and used in conjunction with a digital integrator having a variable triangular window were found to be preferable to RC integrators. The digital integration specified had limited incremental value in increased smoothing of the average EMG signals when the time constants were greater than 35 msec.

Average EMG signals sampled simultaneously from several electrodes placed in a uniformly acting muscle were highly correlated, with an average r = .95. It was found that speaking rate variation affected the time resolution of the average EMG signal at time offsets greater than approximately 1 sec from the line-up point. On the average, however, the individual EMG signals were found to correlate highly with their average EMG signals, r's ranging from .72 to .88. Thus, the averaging technique is successful in producing a reliable, relatively noise-free and undistorted EMG signal, when a brief time window is chosen.

REFERENCES

Basmajian, J. V. (1974) Muscles Alive. (Baltimore: Williams & Wilkins).

Basmajian, J. V. and G. A. Stecko. (1962) A new bipolar indwelling electrode for electromyography. J. Appl. Physiol. 17, 849.

Bell-Berti, F. and H. Hirose. (1975) Palatal activity in voicing distinctions: A simultaneous fiberoptic and electromyographic study. J. Phonetics 3, 69-74.

Bigland, B. and O. C. J. Lippold. (1954) The relation between force, velocity and integrated electrical activity in human muscles. J. Physiol. 123, 214-224.

Bouisset, S. and F. Goubel. (1971) Interdependence of relations between integrated EMG and diverse biomechanical quantities in normal voluntary movements. *Activitas Nervosa Superior 13*, 23-31.

Buchthal, F. (1959) The functional organization of the motor unit: A summary of results. *Am. J. Phys. Med. 38*, 125-128.

Coggshall, J. C. and G. A. Bekey. (1970) EMG-force dynamics in human skeletal muscle. *Med. Biolog. Eng. 8*, 265-270.

Cooper, F. S. (1965) Research techniques and instrumentation: EMG. In *Proceedings of the Conference: Communicative Problems in Cleft Palate*, ASHA Reports *No. 1*, 153-168.

Hirose, H. (1971) Electromyography of the articulatory muscles: Current instrumentation and technique. *Haskins Laboratories Status Report on Speech Research SR-25/26*, 73-85.

Hirose, H. (1976) Posterior-Cricoarytenoid as a speech muscle. *Ann. Otol. Rhinol. Laryng. 85*, 334-343.

Inman, V. T., H. J. Ralston, J. B. de C. M. Saunders, B. Feinstein, and E. W. Wright, Jr. (1952) Relation of human electromyogram to muscular tension. *Electroenceph. Clin. Neurophysiol. 4*, 187-194.

Kewley-Port, D. (1973) Computer processing in EMG signals at Haskins Laboratories. *Haskins Laboratories Status Report on Speech Research SR-33*, 178-183.

Kreifeldt, J. G. (1971) Signal versus noise characteristics of filtered EMG used as a control source. *IEEE Trans. Biomed. Engin. BME-18*, 1, 16-22.

Leifer, L. J. (1969) Characterization of single muscle fiber discharge during voluntary isometric contraction of the biceps brachii muscle in man. Unpublished dissertation, Stanford University.

Lippold, O. C. J. (1952) The relation between integrated action potentials in a human muscle and its isometric tension. *J. Physiol. 117*, 492-499.

McNemar, Q. (1969) *Psychological Statistics*. (New York: Wiley).

Person, R. S. and L. P. Kudina. (1972) Discharge frequency and discharge pattern of human motor units during voluntary contractions of muscle. *Electroenceph. Clin. Neurophysiol. 32*, 471-483.

Trimble, J. L., B. L. Zuber, and S. N. Trimble. (1973) A spectral analysis of single motor unit potentials from human extraocular muscle. *IEEE Trans. Biomed. Engin. BME-March*, 148-151.

Waring, W. (1974) Observing signals from nerve and muscle. In *Biomedical Electrode Technology*, ed. by H. A. Miller and D. C. Harrison. (New York: Academic Press).

Zuniga, E. N. and D. G. Simons. (1969) Nonlinear relationship between averaged electromyogram potential and muscle tension in normal subject. *Arch. Phys. Med. Rehab. 50*, 613-620.

Perceptual Test of a Phonological Rule

Linda Shockey

ABSTRACT

Synthetic speech is characteristically produced in a highly formal, maximally differentiated style. This experiment shows that the output of one phonological rule manifested frequently in conversational English (ð-assimilation) can be simulated easily by lengthening appropriate word-final consonants. Listeners accept the resulting long consonants as consonant + ð clusters. It is suggested that the inclusion of this and other casual speech rules will improve the naturalness and hence the "listenability" of synthetic speech.

It has been noted by Wolfram and Fasold (1974), Gimson (1962), Kohmoto (1965), Hubbell (1950), and Shockey (1974), among others, that in relaxed or conversational speech a word-initial [ð] will assimilate completely to a preceding nasal, fricative, or [l] and will cause the nasal, fricative, or [l] to lengthen as well. Examples from natural speech follow (taken from Shockey (1974):

S # Assimilations:

course they	['kɔɹs·e$^{I}_{\wedge}$]
effects the	[ɨfɛks·ə]

Z # Assimilations:

broads that	[bɹɑdz·æʔ]
cause the	[kɑz·ə]
years there	[ɪ̭ɪɹz·ɛɹ]

N # Assimilations:

seen the	[sin·ə]
on their	[ɑn·ɛɹ]
line that	[la$^{I}_{\wedge}$n·æt]

The ð-assimilation will presumably only occur when the cluster in question receives a relatively low degree of stress. Therefore, if one were to construct a sentence of the sort "I said put it in this box, not in that one!," the assimilation would be less likely to occur. However, since virtually all English words beginning with [ð] carry very little semantic load, ð-assimilation can be expected to be very frequent. This experiment was

---

[HASKINS LABORATORIES: Status Report on Speech Research SR-50 1977]

conducted to investigate whether the lengthening of the assimilating (word-final) consonant is sufficient to induce a cluster percept. If so, ð-assimilation could easily be included in a speech synthesis strategy designed to output natural-sounding English. The specific condition examined here is ubiquitous in conversational speech: in the word "the", [ð] assimilates to a preceding word-final consonant so that the distinction between definite and indefinite articles is preserved mainly in the duration of the final consonant. (It is also possible that another cue may be retained in the consonant transitions. Final alveolar consonants may be fronted and thus be more dental in cases where the interdental [ð] has been assimilated. This possibility must be examined further).

To test the perceptual effect of final consonant lengthening, the following experiment was conducted: two utterances by one female speaker were digitized and stored in a computer-accesssible file. These utterances were "miss a guy" ['mɪsəgaᴵ̭] and "warn a guy" ['ᵘ̭ɑɹnəgaᴵ̭]. Using digital splicing techniques, the duration of the frication in the first utterance was varied from 80 to 200 msec in 20 msec steps, and the duration of the low amplitude, low frequency portion of the acoustic signal that presumably corresponded articulatorily to the closure portion of [n] in the second, varied from 0 to 120 msec in 10 msec steps. The lengthening was done by holding the transitions into and out of the steady-state consonant in question and repeating a characteristic portion of the waveform judged to be the center of the consonant enough times to give the desired durations. It was judged that the nasal closure began when the waveform became smooth and lacked high-frequency components. This decision lies behind the inclusion of an [n] of 0 msec. The heavily nasalized transitions into and out of the closure were sufficient to give an impression of the nasal consonant.

Two tests were constructed. For each test, each stimulus was included four times in a randomized-order listening test with three seconds between stimuli. The resulting tests were presented over headphones to 30 undergraduate students at Ohio University. The students were asked to judge whether the middle word in the three-word sequence was "a" [ə] or "the" [ðə]. They were instructed that the signal had been degraded and, therefore, that their decision was to be based on which English article the stimulus reminded them of the most.

Results are depicted graphically in Figure 1. At the top we see results for the [s] test (miss a guy). It demonstrates that when the [s] assumes a duration of 130 msec, subjects cease to hear the sequence as containing the indefinite article and begin to hear it as containing the definite article.

Figure 1 (bottom) shows the same result at 120 msec for the lengthened nasal segment. We have thus been able to induce the impression of an s + ð cluster or n + ð cluster by increasing the length of the assimilating consonant.

The curves shown in Figure 1 represent data points for all subjects who responded systematically to the stimuli. Of 31 subjects, 5 responded randomly to both tests. Five additional subjects displayed random results for part 2 (nasal + ð assimilation) while performing adequately on part 1. This means that there were 24 subjects for part 1 and 21 for part 2. It is difficult to

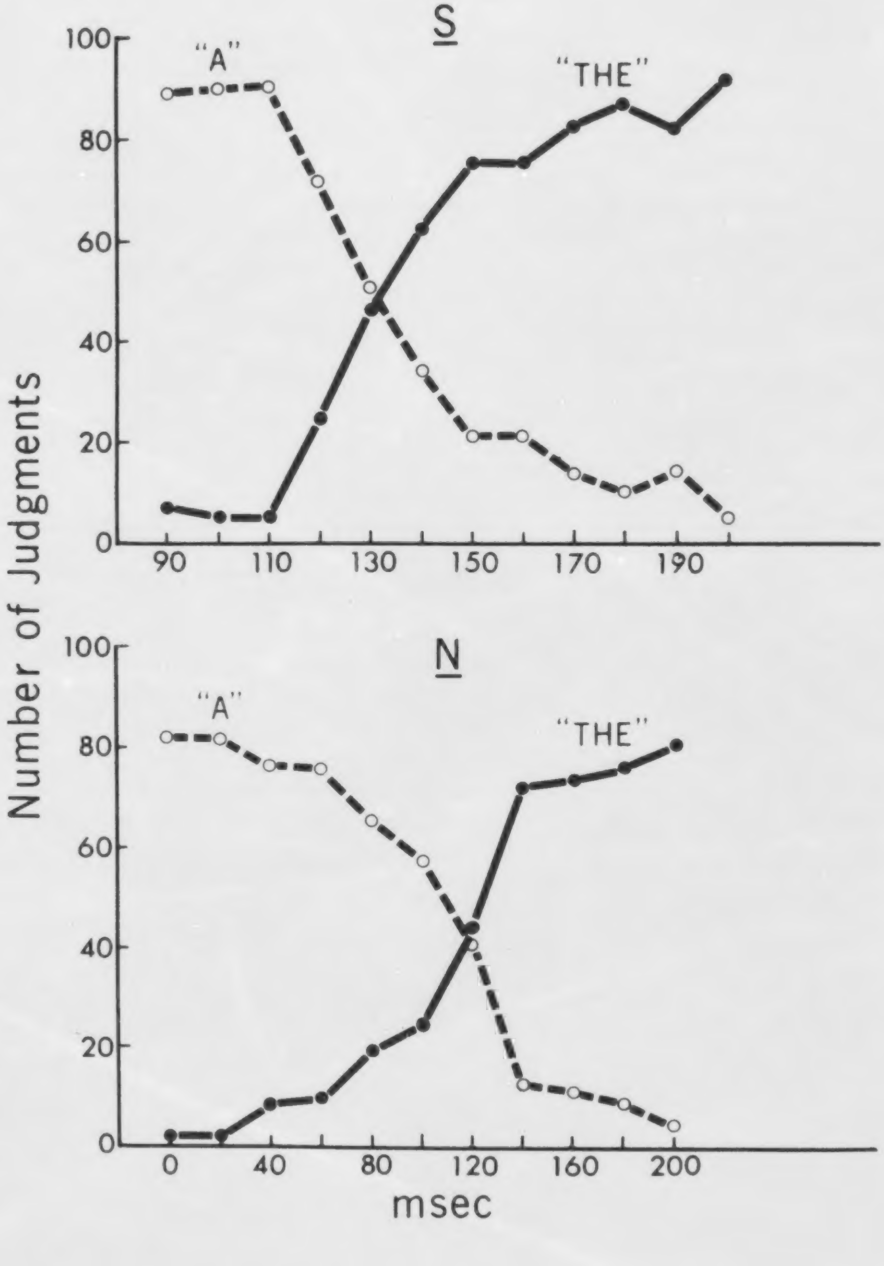


Figure 1: Number of "a" and "the" judgments as the nasal segment in "warn a guy" is lengthened; number of "a" and "the" judgments as the fricative segment in "miss a guy" is lengthened.

say why many people did not respond in a patterned fashion to these stimuli; perhaps the utterance is not long enough to induce a casual speech frame of perception for some. It seems unlikely that the frequency with which one hears these processes is a contributing factor, since both assimilations are extremely common in the dialect areas in which the subjects live (Shockey, 1974).

Practical applications are foreseeable for the results of this experiment:

1) Since [ð] is a difficult sound for learners of English to perfect, they can be taught quite early on to assimilate it to appropriate preceding consonants while also lengthening the consonant. This will not only make their English easier to articulate, but closer to standard conversational speech, since consonant + [ð] clusters are scarce in spoken English. Of course, nonassimilable [ð] must still be dealt with. (Current English-for-non-natives texts regard ð-assimilation as substandard. However, the author has observed it to be ubiquitous in the speech of American newscasters, actors, politicians, and others who depend on effective oral communication.)

2) In speech synthesis strategies, consonant + [ð] clusters can be approximated by lengthening the consonants that participate in this rule, which could make the resulting speech phonologically more natural, less stilted, and easier to listen to for extended periods (as is called for in reading machines for the blind). It is likely that other phonological properties of casual or connected speech are equally easy to simulate and should be included in a synthesis strategy that hopes to approximate natural speech output.

REFERENCES

Gimson, A. C. (1962) An Introduction to the Pronunciation of English. (London: Edward Arnold).

Hubbell, A. F. (1950) The Pronunciation of English in New York City. (New York: King's Crown Press).

Kohmoto, S. (1965) Applied English Phonology. (Tokyo: Tanaka Press).

Shockey, L. (1974) Phonetic and phonological properties of connected speech. Ohio State University Working Papers in Linguistics, Number 17 (May), (Dept. of Linguistics, O.S.U.).

Wolfram, W. A. and Fasold, R. W. (1974) The Study of Social Dialects in American English. (Englewood Cliffs, N.J.: Prentice-Hall, Inc.).

Adaptation of the Category Boundary Between Speech and Non-speech: A Case Against Feature Detectors

Robert E. Remez[†]

ABSTRACT

Two experiments were performed employing acoustic continua that change from speech to nonspeech. The members of one continuum, synthesized on the Pattern Playback, varied in equal steps of change in the bandwidths of the first three formants, from the vowel /ɑ/ to a nonspeech buzz. The other continuum, achieved through digital synthesis, varied in the bandwidths of the first five formants, from /æ/ to buzz. The categorical perception of each continuum was established by standard procedures. Perceptual adaptation on these continua then revealed effects on the category boundaries comparable to those reported for speech sounds. The results are interpreted as suggesting that neither phonetic nor auditory feature detectors are responsible for perceptual adaptation of speech sounds.

INTRODUCTION

In their account of the perceptual process underlying phonetic identification, Eimas and Corbit (1973) combined phonetic feature analysis and hypercomplex cells (Hubel and Wiesel, 1965) to yield phonetic feature detectors, understood as special cortical devices tuned to "listen" to the acoustic stream and extract the phonetic building blocks. This claim seemed reasonable on several counts. First, the elusiveness of the acoustic-phonetic correspondence argued that a higher order evaluation would be needed to accomplish the extraction of the meaning from the acoustic stream (Halle and Stevens, 1962; Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967). Second, a century of efforts directed toward delineating the cortical loci of mental faculties established that speech and language are mediated in a restricted cortical area (Penfield and Roberts, 1959; Geschwind and Levitsky, 1968; Witelson and

---

[†] Also University of Connecticut, Storrs.

Acknowledgment: During the course of this research the author was supported by an NICHD Pre-doctoral Training Grant in Language and Psychology awarded to the University of Connecticut. It is a pleasure to thank Tom Casola, Harriet Greisser, Alvin Liberman, Ignatius Mattingly, Philip Rubin, Michael Studdert-Kennedy, Quentin Summerfield, Michael Turvey, Hal Tzeutschler, and Robert Verbrugge for their aid, comment, and criticism during the development of this study and report. Also, a special thank you goes to Agnes McKeon, who taught me how to paint Playback-style.

[HASKINS LABORATORIES: Status Report on Speech Research SR-50 1977]

Pallie, 1973). Third, the nativist position of the generative transformational enterprise argued that rather detailed linguistic knowledge might be prewired into every human infant (Chomsky, 1965; Chomsky, 1968). Fourth, perceptual experiments on neonates seemed to show that infants were sensitive to some phonetic distinctions before any relevant experience (Eimas, Siqueland, Jusczyk, and Vigorito, 1971). And fifth, developments in electrophysiology suggested that the sensitivities of single cortical cells are both elaborate and correlated with the ecology of the animals studied (see Kuffler, 1973, for a review). The notion of genetically pretuned hypercomplex cells that translate auditory events into phonetic descriptions neatly addressed these issues.

Perceptual adaptation of phonetic category boundaries has been the technique by which the properties of the hypothesized detectors have been examined[1], but this body of research only marginally confirms the original detector description. Although certain adaptation effects have required the explanation to include a phonetic level of analysis at which particular acoustic-auditory values are less perceptually significant (Ades, 1974; Diehl, 1975; Miller, 1975; Remez, Cutting, and Studdert-Kennedy[2]), other research has produced evidence of non-phonetic adaptation that is fully compatible with any of the previously obtained phonetic effects (Ades, 1974; Pisoni and Tash, 1975; Bailey, 1975; Diehl, 1976). In short, the present situation is paradoxical. While the underlying detectors are phonetic by the original intention as well as by occasional necessity, some of them may suffer acoustic fatigue, and all suffer from inexplicit specification of their tuning curves. [In tonotopically organized cells, the dimension for measuring sensitivity is frequency (Woolsey and Walzl, 1942); in phonetically organized detectors, the dimensions of analysis must correspond to those of vocal production, and many of these have yet to be defined.] Additionally, and perhaps fatally, the passive filtration method of phonetic feature extraction in speech assumes a simple relation between the acoustic pattern, the phonetic segments, and the ordinal correspondence between them. But it is the fundamental point of many perceptual studies that segmental identity typically is carried by the sound pattern distributed across the entire syllable (for example, Cooper, Delattre, Liberman, Borst, and Gerstman, 1952). This requires, in essence, that each feature detector be a little homunculus, omniscient on the nature of the context conditioned variation of its favorite feature of speech. For these reasons, either the task requires vast multiplexing, well beyond the sensible (Halwes and Jenkins, 1971), or the loss of the appealing simplicity of the device in order to deal with the complexity of the message structure of the speech chain.

---

[1] Haggard (1967) first described the paradigm of after-effect research employing speech sounds, but his rationale was completely independent of neurophysiological claims.

[2] Remez, R. E., Cutting, J. E. and Studdert-Kennedy, M. Acoustic similarity or phonetic identity: a cross adaptation study employing song and string. Paper in preparation.

A test of the hypothesis that adaptation effects reveal detectors tuned to linguistic features might attempt to produce adaptation outside the set of effects predictable from the feature inventory. The experiments reported here, which use acoustic continua from speech to nonspeech sounds, satisfy this condition. Further, if acoustic-auditory explanations can be ruled out, there would then be reason to suggest an active, cognitive basis for this effect, one which, if applied to speech, would be compatible with a phonetic level of analysis, but incompatible with phonetic feature detectors.

EXPERIMENT *Ia*

Methods

Subjects. Sixteen University of Connecticut undergraduates, whose participation fulfilled the introductory psychology course requirement, served as listeners in this part of the study. All were native English speakers with no known speech or hearing disorder or psychopathology. None had any experience with synthetic speech sounds before the listening session.

Stimuli. The Haskins Laboratories Pattern Playback (Cooper, Liberman, and Borst, 1951) was used to synthesize the basic materials[3]. This device uses a tone wheel to generate the harmonics of 120Hz in light intensities arrayed in a frequency scale. A graphic pattern selectively reflects portions of this scale, and this reflection, through capture by a photocell, is transduced to a frequency by amplitude by time acoustic signal. Figure 1 displays the pattern painted (Liquitex Acrylic Titanium White Grumbacher #4) on the acetate belt (Eastman Kodak) and the frequency values of the transduced signal. The pattern changes from a vowel /ɑ/, with formant frequency values of 600 Hz, 1200 Hz, and 2400 Hz, to a nonspeech buzz, by modifying the bandwidths of the formants; initially, the bandwidths are 100 Hz, and they increase to effectively infinite width at the end of the pattern. Figure 2a presents one spectral section through each of the endpoints.

This 1-second sound was transferred to audiotape and then digitized by the Haskins Pulse-Code Modulator (PCM) (Cooper and Mattingly, 1969), sampling at 10 KHz with low-pass filtering at 5 KHz. A ten-step continuum was then made by editing the digitized waveform. Nine cuts were made, one every 100 milliseconds; the oscillographic patterns were equated for amplitude, producing 10 tokens, each of 12 pitch periods which vary in formant bandwidth as does the overall pattern.

Two test tapes were created using the PCM system. An identification sequence contained 10 occurrences of each of the 10 continuum items, for a test of 100 trials, with 5 seconds between trials, and 9 seconds following each decade. A discrimination sequence consisted of ABX triads with 1 second

---

[3] Neither the Haskins Parallel Resonance Synthesizer nor the Ove III were suitable for this study because of hardware-imposed limits on formant bandwidth. These devices are devoted speech synthesizers, and this study required a full-frequency synthesizer, that is, one with no such restriction.

FIGURE 1

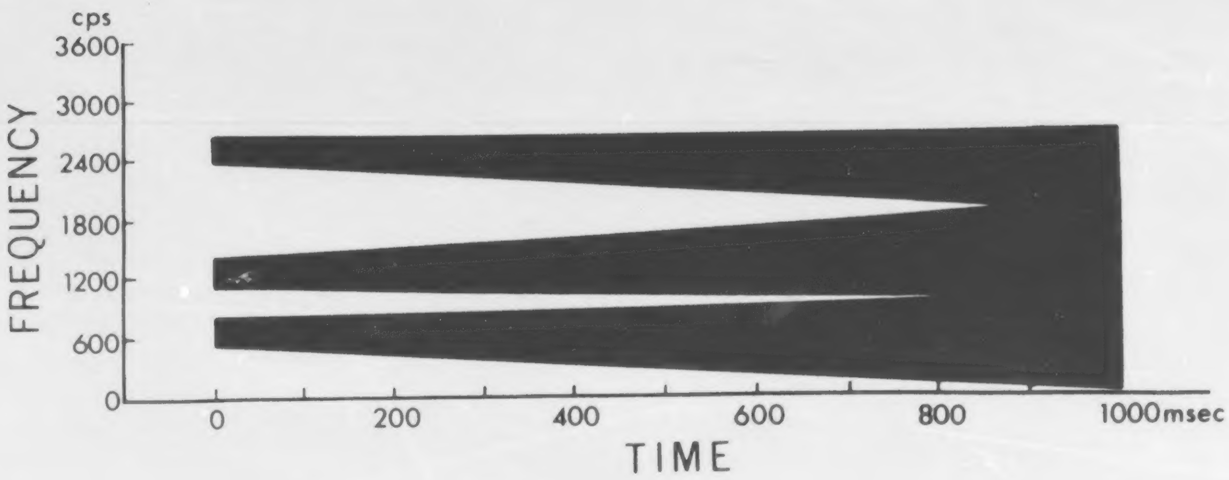


Figure 1: The painted pattern which the Playback reproduced.

FIGURE 2

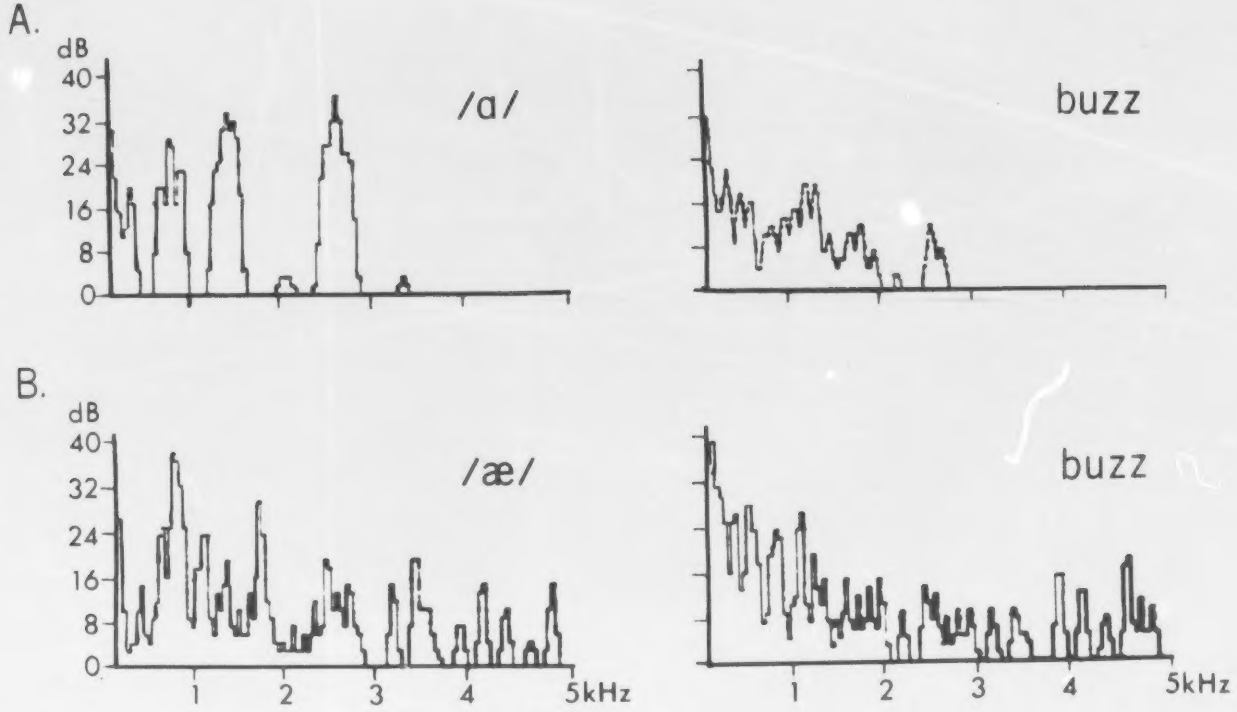


Figure 2: Spectral sections through the endpoints of the Playback-synthesized continuum (a), and the software synthesized continuum (b).

between items, 5 seconds between trials, and 9 seconds separating the decades. The 4 permutations of each ABX comparison were represented: ABA, ABB, BAB, and BAA. In a one-step discrimination, the comparisons are items 1 and 2, 2 and 3, 3 and 4, and so on; at 4 trials per comparison, and 9 comparisons, there were 36 trials in this test.

Procedure and Apparatus

The sixteen listeners were tested in four groups of four. Sounds were presented binaurally over Grason-Stadler earphones activated by a Crown 820-144 tape recorder through a junction box so that several listeners could listen simultaneously. Each session commenced with a briefing sequence in which the endpoints of the continuum were each repeated ten times, in alternation. At that time, listeners were asked to signify if they had a good idea of the sounds' identity; their instructions were to consider the buzz a machine noise, and the vowel a synthetic speech sound. The identification test was then administered. Identifications were scored on a response sheet as speech or buzz (S or B). After a short intermission, listeners were given sample ABX sequences in which they judged which of the first two sounds was identical to the third; the continuum endpoints were used here to insure clarity of the instructions. The actual test, begun when all agreed that they understood the instructions, consisted of the 36-trial discrimination sequence played twice.

Results and Discussion

Three subjects were dropped because they either failed to follow instructions (two subjects declined to judge difficult items) or responded at chance on the identifications (one subject). Results for identification and discrimination appear in Figure 3a. Each point is the mean of 130 observations in the identification test, and 108 observations in the discrimination test. These functions are reasonably consistent with the criteria for categorical perception (Studdert-Kennedy, Liberman, Harris, and Cooper, 1970), in that a peak in discriminability occurs at the breakpoint between the identification categories. The term "categorical perception" describes a situation in which the judged difference between two entities is contingent on their identities rather than on the physical differences between them.

One anomaly in the discrimination function should be addressed, namely, the troughs of the function. Given two categories, speech and buzz, the peaks should number one, not three. In the function of Figure 3, however, the discrimination peaks between items 1 and 2, and between 9 and 10, are as prominent as the category boundary peak. Examination of the waveforms of these tokens revealed amplitude differences between 2 of the 12 pitch periods in each item of the pairs. No such difference could be discovered between items in the other pairs in the set. One possible account, then, of the spurious peaks is that they result from judgments of amplitude rather than spectrum differences. If this reasoning is correct, the peaks can be discounted in any challenge to categoricity, because the manipulation of interest is spectral, and this particular discrimination is made on a nonspectral basis.

FIGURE 3

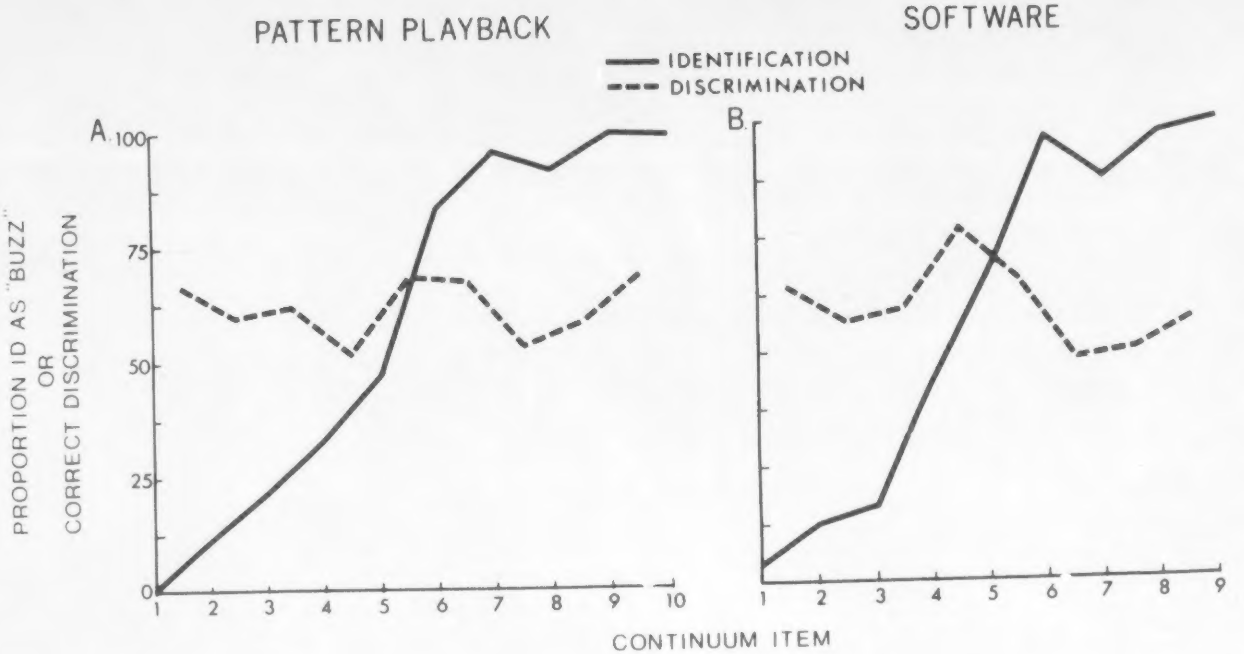


Figure 3: Identification and discrimination plots for Playback (a) and software (b) continua.

EXPERIMENT Ib

Methods

Subjects. Eight University of Connecticut undergraduates were paid to listen in this part of the study. They had all participated in Experiment Ia (five from the original group could not attend the listening sessions for scheduling reasons).

Stimuli. The 10 tokens from Experiment Ia were used. An adaptation sequence consisted of an initial 100 repetitions of the adapting item, one of the continuum endpoints, at 1-second intervals. After a 10-second pause, which cued the listeners that the identification trials were coming up, six items from the continuum were presented for identification, as either speech or buzz (S or B). At the conclusion of the block of six, there was a 10-second pause followed by 50 repetitions of the adapting item, another block of six, and so on for the remainder of the test.

Each of the 10 sounds drawn from the continuum was presented for identification twelve times, with the exception of the four most extreme, the two on each end that were presented six times each. This preserves sensitivity in the midrange of the continuum and shortens the test by two blocks of identifications, to the relief of the listener. With 96 trials (6 twelves and 4 sixes) there were 19 blocks of six trials each. The random order for these items was the same in both the speech and the buzz adaptation sequences.

Procedure. An identification sequence was used to determine a standard identification function in each of the two sessions. This was used for comparison with the adapted identification function. All subjects took part in both conditions; half took the /ɑ/ test first; half took the buzz test first. Several days separated the test sessions. The equipment and test conditions were in all other respects the same as in Experiment Ia.

Results

Each subject contributed two sets of judgments per session, a pretest set and an adaptation set. To each of these a standard ogive was fitted, after Woodworth (1938). Thus, two scores were available for each subject per test, one mean of the fitted ogive, measured in continuum units, for the pretest, and one for the adaptation test.

The curves for the grouped data for each session appear in Figure 4. Each pretest plot represents the means of 80 trials per continuum item; in the adaptation plot, the two extreme points on either end, items 1, 2, 9, and 10, are the means of 48 observations each; the remaining six medial points are the means of 96 judgments each. The change in the ogive mean due to speech adaptation was 1.32 continuum units, and that due to buzz was .638 continuum units in the opposite direction.

A two-factor repeated measures analysis of variance was performed on the ogive means, with two levels of adaptor (Speech or Buzz) and two conditions (Pretest or Adaptation). There were no significant main effects, indicating

FIGURE 4

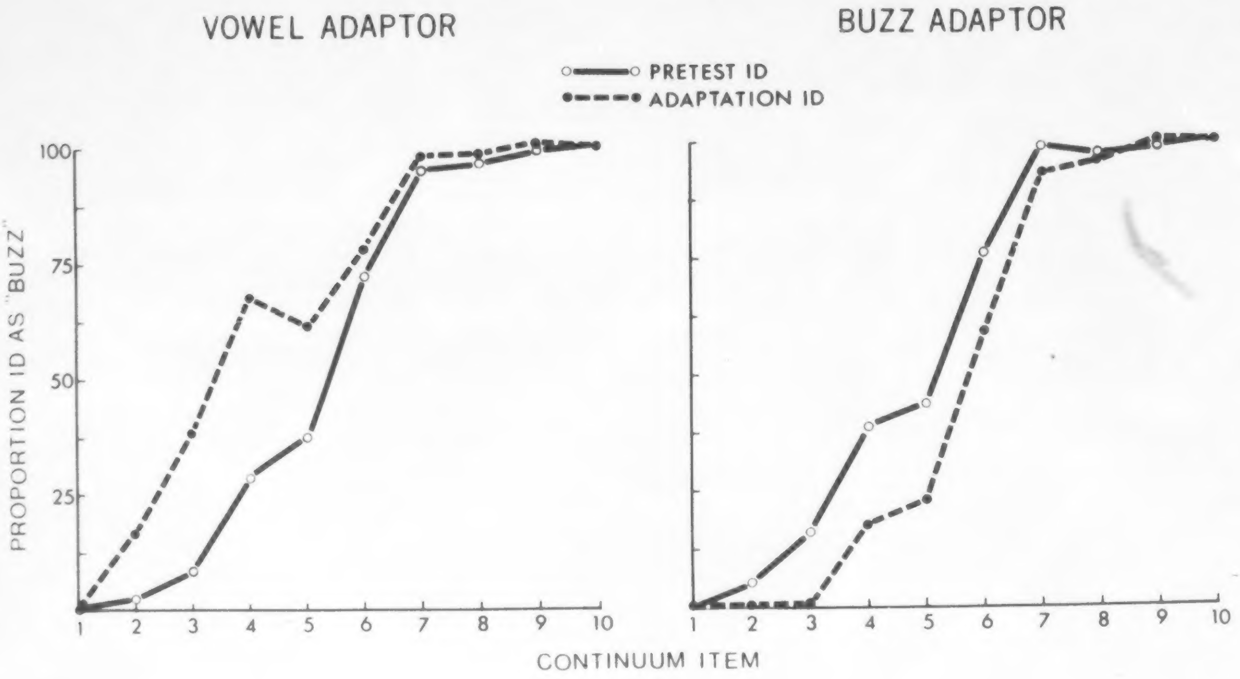


Figure 4: Pretest and adaptation test curves for the Playback continuum.

that the means, when collapsed across conditions or across adaptor, did not depart from the grand mean. However, the interaction of adaptor by condition was significant [$F(1,7)=12.771$, $p < .01$)], reflecting the different, and opposite, effect of each adaptor relative to the pretest mean, on the adaptation test mean.

Discussion

In this experiment the shifts of the category boundary do damage to the proposal that the units mediating the effect are isomorphic with the primitives of phonetic feature analysis. Miller (1975), for example, has proposed that feature analyzing devices are arranged so that the fatigue of one leads to the relatively enhanced strength of its inverse or opponent. However, in the case of /ɑ/ fatigue, the resulting change requires going outside the feature set used in phonetic descriptions to capture the distinction between speech and nonspeech. In other words, a processing device arranged along the lines of phonetic features could neither predict nor explain this case of adaptation[4].

Must we then have recourse to a "lower level" account of the results? Other authors have found higher order descriptions unwarranted by the adaptation effects (Pisoni and Tash, 1975; Bailey, 1975; Ades[5]), and since higher order phenomena can be described in lower order physical terms, they have ascribed their effects to alterations in sensitivity at a lower level within the auditory system. For example, a receptive unit with a "best" stimulus may tend to show, under fatigue, a decreased sensitivity to that stimulus; its decrease leads to the release from inhibition of adjacent, similar receptors, and consequently, to the relative enhancement of near misses from the best value. Thus, the auditory system, early on in the course of an analysis, can yield a mistransformed description which the unsuspecting analyzers in the next step are helpless to reverse. [The actual neurophysiology of this is still open to question. One current topic of investigation is whether the reciprocal inhibition demonstrable at the VIII nerve nucleus arises cochlearly or in the nucleus itself (Mountcastle, 1974).] By this mechanism, then, a unit or units sensitive to a portion of the frequency range, when fatigued, will be less sensitive to the absolute values of stimulation, and will, via disinhibition, effectively amplify departures from the original fatigued values. For example, a receptor that mediates a rising second formant value over the course of 35 msec (which specifies a voiced bilabial in some circumstances) will be less sensitive when fatigued to values that exactly conform to the pattern of fatigue. Disinhibition of neighboring receptors would, in effect, boost receptor responses to second formant transitions that depart by small amounts from the fatigued values. This will be costly in terms of the perceptual outcome only in borderline cases, that is, those in which the fatigue-disinhibition throws the pattern over the line from one category into another. This explanation insists that the auditory transcription which the phonetic system is given to work with has been irretrievably altered.

---

[4] Morse, Kass, and Turkienicz (1976) found that an /ɛ/ or /i/ adaptor, but not /ɪ/, changed both boundaries on an /i/-/ɪ/-/ɛ/ continuum; they concluded that the binary distinctions "tense" and "high" of Chomsky and Halle (1968), which would specifically rule out such a finding, had been empirically falsified. They offered that, because their results had shown continuity rather than discreteness, the feature system underlying the result was necessarily continuous, many-valued rather than binary; by extrapolation, so was the detector. Their approach was like that of Cooper and Blumstein (1974) who were the first to use adaptation to find perceptual interactions and thereby to define the phonetic features perceptually, rather than acoustically or articulatorily. Nevertheless, because no language makes phonetic use of the distinction [+speech, -speech], it is safe to say that /ɑ/-buzz adaptation does not require that we posit a new feature; rather, it undermines the interpretation of speech adaptation effects in terms of phonetic features.

[5] Ades, A. E. Source assignment and feature extraction in speech. (unpublished manuscript).

However, if this reasoning is applied to the adaptation by /ɑ/ and buzz, a curious situation arises. Fatigue caused by buzz should decrease sensitivity throughout the range of frequencies used; no frequency-specific effects should occur. Indeed, the auditory view of adaptation would predict no adaptation at all.

On the other hand, fatigue caused by /ɑ/ should eat holes in the "neural spectrogram" of the buzz, decreasing sensitivity at 600 Hz, 1200 Hz, and 2400 Hz. If a listener were then presented with the buzz, he should hear a pattern the inverse of /ɑ/, with formants at 300 Hz, 900 Hz, and 1800 Hz[6]. In this case, if the listener judges the sound on the basis of the acoustic feature of presence or absence of formant structure, then the auditory point of view predicts that the boundary should move toward the buzz, since the fatigued spectral receptors, even in this extreme case, might be expected to retain a pattern showing acoustic maxima and minima. However, precisely the reverse boundary movement was actually observed.

The present experiment, therefore, produces a situation unique in the adaptation literature. While the conventional approach has been to suspect auditory processes by default whenever a phonetic account fails, this is obviously not possible here. The listener must be judging the sounds on other than a simple acoustic basis. The perception of the novel distinction between /ɑ/ and buzz required here indicates a stable perceptual capacity that can be reconfigured to suit the demands of the particular situation. Neither phonetic feature nor acoustic property detectors can be reconciled with this type of perceptual flexibility.

---

[6] This pattern, when rendered by the Playback, does not sound like a speech sound.

Based on the foregoing, there are several motives for extending this investigation. First, the anomalous discrimination peaks mildly threaten the claim of categoricity for this perceptual distinction. By implication, the explanation that the adaptation effect is judgmental rather than perceptual, cannot be confidently rejected. Second, the artificiality of the speech synthesized on the Playback may be a factor to consider. Because the Playback, however phonetically identifiable its message may be, has a voice quality unlike that of any person, its use in an experiment of this kind may produce synthesizer artifacts. As a precaution, then, it would be valuable to try this procedure with a more natural sounding synthesizer. Third, the possibility that this effect is restricted to /ɑ/, that /ɑ/ might intrinsically be more nonspeechlike than other vowels, could be assessed by using a different vowel in the same paradigm. On these accounts, Experiment II was performed.

EXPERIMENT IIa

Methods

Subjects. Eight University of Connecticut undergraduates, not those of Experiment I, were paid for their participation. All were naive with respect to synthetic speech.

Stimuli. The software synthesizer of Fisher and Engebretson (1975), modified to permit variable parameterization of all five formants, was used to make the acoustic tokens. This program calculates a digital wave from user-determined parameters of source frequency, formant frequency and bandwidth, and overall duration and amplitude. The digital wave is then converted to audio via a PCM system. These programs, implemented by Joe Kupin and Hal Tzeutschler, run on the University of Connecticut Language and Psychology Data General NOVA 2.

A nine-step continuum from /æ/ to buzz was made by successive 50 Hz increments in the formant bandwidths, starting from an initial bandwidth of 100 Hz for each formant. Duration was 140 msec; overall amplitude was 45 dB; fundamental frequency was 120 Hz; formant frequencies for the vowel were F_1:750, F_2:1650, F_3:2460, F_4:3500, and F_5:4500. The audio output was transferred to the Haskins PCM via Ampex tape recording, to permit algorithmic envelope shaping. Each item was sixteen pitch periods (140 msec) long, with ramp on and off of 3 periods (25 msec); overall amplitudes were equated. Spectral sections through the endpoints appear in Figure 2b.

The identification test consisted of ten judgments of each of the nine items in random order. The discrimination test consisted of eight judgments of each of the eight one-step comparisons, in random order.

Procedure and apparatus. The outline of Experiment Ia was followed.

Results and Discussion

Figure 3b displays the functions for identification and discrimination. The identification plot displays the means of 80 trials per point, the discrimination plot the means of 64 trials per point. Inspection of the figure will reveal that, relative to Experiment Ia, the peak at the category boundary remains a property of the discrimination function, while the peaks at the extremes of the continuum have disappeared at the buzz end, and all but disappeared at the speech end. It is reasonable to conclude that this more carefully controlled continuum elicited true categorical perception.

EXPERIMENT IIb

Methods

Subjects. The eight listeners from IIa were paid for their participation in this section of the study.

Stimuli. The nine-item continuum from IIa was used to make the adaptation sequences. These tests differed from Ib only in the consequences of using a nine- as opposed to a ten-step continuum. Here, the four most extreme items were presented six times each for identification during adaptation, and the remaining five medial items twelve times each. With 84 trials overall (4 sixes and 5 twelves), there were fourteen blocks of six trials each, which alternated with the repeating adaptation item, either /æ/ or buzz. The random order of identifications during adaptation was the same in the speech and buzz adaptor conditions.

Procedure. As in the previous procedure, each test day began with the identification sequence in order to obtain a standard for comparison with the adapted identification; test days were consecutive. All subjects took part in both adaptor conditions that were counterbalanced for order.

Results

The ogive fitting method was again used on the two tests per day contributed by each subject.

Averaged functions for both adaptation conditions appear in Figure 5. Pretest plots show the means of 80 trials per continuum item; adaptation plots show the means of 48 trials for items 1, 2, 8, and 9 and 96 trials for 3 through 7. The change in the ogive mean due to speech adaptation is .883 continuum units, due to buzz adaptation .692 in the opposite direction.

An analysis of variance was performed on the ogive means, with two levels of each factor, adaptor (Speech/Buzz) and mean (Pre/Post). The interaction is the term of interest here; with $F(1,7)=32.842$, $p < .001$. The main effect of adaptor was also significant; with $F(1,7)=65.858$, $p < .001$. The statistical significance of the adaptor term was due to the close correspondence of the two pretest means, which, when averaged with the adaptation means, clearly reflect the differential effects of adaptation. (Experiment Ib showed no such significance for this term because the pretest means varied in opposition to the adaptation means, thus cancelling the effect of adaptor upon averaging.)

FIGURE 5

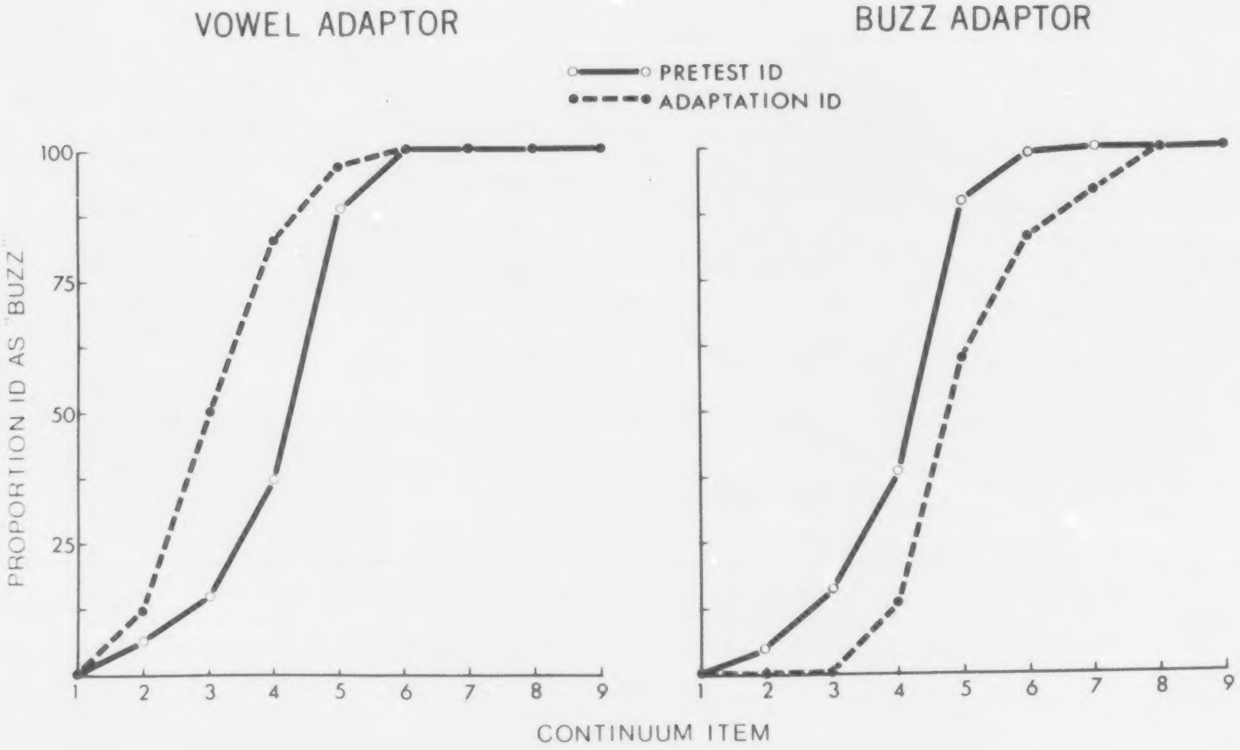


Figure 5: Pretest and adaptation test curves for the software continuum.

Discussion

This study with software synthesized sounds strengthens the original argument made from Playback data. The results show that it is neither the artificiality of the speech synthesizer nor the particular vowel involved that enables listeners to treat the present acoustic continua in the same fashion as continua of proper speech sounds. It is this comparability which suggests the interpretation that a speech/nonspeech detector may be responsible here, a detector just like those that putatively underlie phonetic adaptation. However, an auditory feature account is ruled out by the argument presented earlier, leaving either (1) a phonetic-type detector or (2) a detectorless explanation as the readily apparent alternatives to consider.

A phonetic detector explanation here would require the extension of the detector inventory, since a speech/nonspeech feature is not found in linguistic analysis. The distinction, in fact, is not even truly linguistic, in the sense of distinctive feature theory, but it certainly is a feature of human perceptual sensitivity, and on that basis might be seen as potentially detector-mediated. However, the existence of perceptual sensitivity should not be the only criterion for postulating a detector. The very advantage of this style of pattern recognition is that it makes infinite use of finite means; if a new detector is to be added to the set at every new discovery, then the contradiction of an indefinitely expandable finite means reduces the attractiveness of the model. The device required by these data can preserve its economy only if it has a small group of detectors, tuned to speech, set in opposition to a small group tuned to nonspeech. On this account, the search for independent confirmation of this organizational plan does not find the neurophysiology encouraging. Although there have been discussions of single-cell mediation of all perception, along the lines of innate taxa (Stent, 1975), as well as descriptions of arrays of phonetic single units (Miller, 1975), no proposal has yet been made to oppose speech neurons and nonspeech neurons. In fact, some claims for uniqueness of the speech neurology imply that the speech processor, whatever it may be, is separate from the nonspeech processor (Milner, 1962). Speech, in this view, is a mode, like vision or audition, and, by analogy, interacts with other modes but is independent of them. In short, a vast opponent process system for speech/nonspeech is not to be endorsed on the basis of any current view, and it may be presumed, in addition, that such a system is unlikely to exist given what is already known.

Finally, the only direct evidence for feature detectors in speech, as opposed to the invitation to such a conceptualization offered by neurophysiological metaphor, is the selective adaptation work. Boundary shifts occasioned by adaptation are precisely the effects that would permit the perceptual correlates of phonetic feature manipulations to be recast as the products of hypothetical detectors. However, though the hypothesis is reasonable when the endpoints differ by a single feature, it is difficult to imagine that a vowel and a buzz are also distinguished by but a single feature, speech/non-speech. The adaptation technique, the only test for feature detectors, is, ironically, not a demonstration of feature detectors at all. It simply reveals that certain perceptual contrasts, in this particular case of higher order properties, undergo selective alteration following saturation.

This study of vowel-buzz adaptation suggests that because the hypothetical detectors are incapable of handling the result, and because the detectors required to handle it are implausible, selective adaptation does not depend on the existence of feature detectors. If the basis for adaptation, and perhaps speech perception as well, can be understood as sensitivity to the higher-order values inherent in acoustic pressure fluctuations, without decomposition into features, then the description of such a process, not mere verification of analytic features, is the goal toward which further research might well proceed.

REFERENCES

Ades, A. E. (1974) How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Percept. Psychophys. 16*, 61-66.

Bailey, P. J. (1975) Perceptual adaptation in speech. Unpublished dissertation, Cambridge University.

Chomsky, N. (1965) *Aspects of a Theory of Syntax*. (Cambridge: MIT).

Chomsky, N. (1968) *Language and Mind.* (New York: Harcourt, Brace and World).

Chomsky, N. and M. Halle. (1968) *The Sound Pattern of English*. (New York: Harper & Row).

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (1952) Some experiments on the perception of synthetic speech sounds. *J. Acoust. Soc. Am. 24*, 597-606.

Cooper, F. S., Liberman, A. M., and Borst, J. M. (1951) The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proc. Natl. Acad. Sci. 37*, 318-325.

Cooper, F. S. and Mattingly, I. G. (1969) Computer controlled PCM system for investigation of dichotic speech perception. *J. Acoust. Soc. Am. 46*, 115(A).

Cooper W., and Blumstein, S. (1974) A "labial" feature analyzer in speech perception. *Percept. Psychophys. 15*, 591-600.

Diehl, R. (1975) The effect of selective adaptation on the identification of speech sounds. *Percept. Psychophys. 17*, 48-52.

Diehl, R. (1976) Feature analyzers for the phonetic dimension stop vs. continuant. *Percept. Psychophys. 19*, 267-272.

Eimas, P. D. and Corbit, J. D. (1973) Selective adaptation of linguistic feature detectors. *Cog. Psychol. 4*, 99-109.

Eimas, P. D., Siqueland, E. P., Jusczyk, P., and Vigorito, J. (1971) Speech perception in infants. *Sci. 171*, 303-306.

Fisher, W. M. and Engebretson, A. M. (1975) Simple digital speech synthesis. *Am. J. Computational Linguistics*, Microfiche 16.

Geschwind, N. and Levitsky, W. (1968) Human brain: left-right asymmetries in temporal speech region. *Sci. 161*, 186-187.

Haggard, M. (1967) Figural after-effect phenomena for speech and non-speech stimuli. *Haskins Laboratories Status Report on Speech Research SR-9*, 61-62.

Halle, M. and Stevens, K. (1962) Speech recognition: a model and a program for research. *IRE Trans. Inform. Theory* IT-8(2), 155-159.

Halwes, T. and Jenkins, J. J. (1971) Problem of serial order in behavior is not resolved by context sensitive associative memory models.

Psychol. Rev. 78, 122-129.

Hubel, D. H. and Wiesel, T. N. (1965) Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *J. Neurophysiol. 28*, 229-289.

Kuffler, S. W. (1973) The single cell approach in the visual system and the study of receptive fields. *Invest. Opthalmol. 12*, 794-813.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychol. Rev. 74*, 431-461.

Miller, J. L. (1975) Properties of feature detectors for speech: Evidence from the effects of selective adaptation on dichotic listening. *Percept. Psychophys. 18*, 389-397.

Milner, B. (1962) Lateralization effects in audition. In *Interhemispheric Relations and Cerebral Dominance*, ed. by V. B. Mountcastle. (Baltimore: Johns Hopkins Press).

Morse, P. A., Kass, J. E., and Turkienicz, R. (1976) Selective adaptation of vowels. *Percept. Psychophys. 19*, 137-143.

Mountcastle, V. B. (1974) Central neural mechanisms in hearing. In *Medical physiology*, vol. I, ed. by V. B. Mountcastle. (St. Louis: C. V. Mosby).

Penfield, W. and Roberts, L. (1959) *Speech and Brain Mechanisms*. (Princeton: Princeton U. Press).

Pisoni, D. B. and Tash, J. (1975) Auditory property detectors and processing place features in stop consonants. *Percept. Psychophys. 18*, 401-408.

Stent, G. S. (1975) Limits to the scientific understanding of man. *Sci. 187*, 1052-1057.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., and Cooper, F. S. (1970) Motor theory of speech perception: a reply to Lane's critical review. *Psychol. Rev. 77*, 234-249, 1970.

Witelson, S. and Pallie, A. (1973) Left hemisphere specialization for language in newborn. *Brain 96*, 641-646.

Woodworth, R. S. (1938) *Experimental Psychology*. (New York: Holt).

Woolsey, C. N. and Walzl, E. M. (1942) Topical projection of nerve fibers from local regions of the cochlea to the cerebral cortex of the cat. *Bull. Johns Hopkins Hosp. 71*, 315-344.

II. CUMULATIVE INDEX

CUMULATIVE INDEX

STATUS REPORT VOLUME NUMBERS

1965: SR-1, SR-2, SR-3, SR-4
1966: SR-5/6, SR-7/8
1967: SR-9, SR-10, SR-11, SR-12
1968: SR-13/14, SR-15/16
1969: SR-17/18, SR-19/20
1970: SR-21/22, SR-23, SR-24
1971: SR-25/26, SR-27, SR-28
1972: SR-29/30, SR-31/32
1973: SR-33, SR-34, SR-35/36
1974: SR-37/38, SR-39/40
1975: SR-41, SR-42/43, SR-44
1976: SR-45/46, SR-47, SR-48
1977: SR-49, SR-50

SUPPLEMENTS

1966: Singer (October)
1967: Gay (June)
1968: Mattingly (April)
1969: Halwes (September)
1970: Kirstein (Bound with SR-24)
1971: Raphael (August), Pisoni (November)
1973: Bell-Berti (September), Wood (Bound with SR-35/36)

CUMULATIVE INDEX

1965-1971 (SR-1 - SR-28): SR-28
1972-1973 (SR-29/30 - SR-35/36): SR-35/36
1965-1976 (SR-1 - SR-49): SR-50

AUTHOR INDEX

Haskins Laboratories Status Report on Speech Research

SR-1 (1965) - SR-49 (1977)

Abramson, Arthur S. (See Lisker, SR-1, 3.1)

__________ (See Hadding-Koch, SR-2, 2.1)

__________ and L. Lisker. Voice Onset Time in Stop Consonants: Acoustic Analysis and Synthesis. SR-3 (1965), 1.1.

__________ Book Review: Mary R. Haas, *Thai-English Student's Dictionary*. SR-3 (1965), 4.1.

__________, L. Lisker and F. S. Cooper. Laryngeal Activity in Stop Consonants. SR-4 (1965), 6.1.

__________ (See Lisker, SR-5/6, 1.1)

__________ (See Lisker, SR-5/6, 4.1)

__________ On Matching the Speech Signal to Mechanisms of Production. SR-10 (1967), 2.1.

__________ (See Lisker, SR-11, 9)

__________ and L. Lisker. Discriminability Along the Voicing Continuum: Cross-Language Tests. SR-11 (1967), 17.

__________ and L. Lisker. Laryngeal Behavior, the Speech Signal, and Phonological Simplicity. SR-11 (1967), 23.

__________ and L. Lisker. Voice Timing: Cross-Language Experiments in Identification and Discrimination. SR-13/14 (1968), 49.

__________ (See Lisker, SR-15/16, 81)

__________ Tonal Experiments in Whispered Thai. SR-19/20 (1969), 37.

__________ (See Sawashima, SR-19/20, 201)

__________ (See Lisker, SR-21/22, 201)

__________ Phonetics: An Overview. SR-23 (1970), 3.

__________ (See Lisker, SR-27, 133)

__________ (See Lisker, SR-27, 175)

__________ and L. Lisker. Voice Timing in Korean Stops. SR-27 (1971), 179.

__________ Word-Final Stops in Thai. SR-28 (1971), 193.

__________ and L. Lisker. Voice-Timing Perception in Spanish Word-Initial Stops. SR-29/30 (1972), 15.

__________ (See Hirose, SR-31/32, 183)

__________ (See Erickson, SR-31/32, 231)

__________ The Tones of Central Thai: Some Perceptual Experiments. SR-39/40 (1974), 91.

__________ Pitch in the Perception of Voicing States in Thai: Diachronic Implications. SR-41 (1975), 165.

__________ The Coarticulation of Tones: An Acoustic Study of Thai. SR-44 (1975), 119.

__________ Thai Tones as a Reference System. SR-44 (1975), 127.

__________ Laryngeal Timing in Consonant Distinctions. SR-47 (1976), 105.

__________ Static and Dynamic Acoustic Cues in Distinctive Tones. SR-47 (1976), 121.

Atkinson, James E. Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency. (Abstract) SR-34 (1973), 211.

__________ (See Erickson, SR-45/46, 205)

__________ and D. Erickson. The Function of Strap Muscles in Speech: Pitch Lowering or Jaw Opening? SR-49 (1977), 97.

Baer, Thomas, T. Gay, and S. Niimi. Control of Fundamental Frequency, Intensity, and Register of Phonation. SR-45/46 (1976), 175.

Bailin, Howard M. (See Borst, SR-3, 2.1)

Bartlett, James C. (See Day, SR-28, 93)

__________ (See Day, SR-31/32, 49)

Bell-Berti, Fredericka and H. Hirose. Stop Consonant Voicing and Pharyngeal Cavity Size. SR-31/32 (1972), 207.

________ and H. Hirose. Velar Activity in Voicing Distinctions: A Simultaneous Fiberoptic and Electromyographic Study. SR-31/32 (1972), 223.

________ (See Raphael, SR-33, 203)

________ and H. Hirose. Patterns of Palatoglossus Activity and their Implications for Speech Organization. SR-34 (1973), 203.

________ The Velopharyngeal Mechanism: An Electromyographic Study. (Ph.D. dissertation) Supplement to the Status Report, September 1973.

________ and K. S. Harris. The Motor Organization of Some Speech Gestures. SR-35/36 (1973), 1.

________ and K. S. Harris. More on the Motor Organization of Speech Gestures. SR-37/38 (1974), 73.

________ and K. S. Harris. Some Acoustic Measures of Anticipatory and Carryover Coarticulation. SR-42/43 (1975), 297.

________ and K. S. Harris. Some Aspects of Coarticulation. SR-45/46 (1976), 197.

Benson, Peter. (See Preston, SR-13/14, 179)

Berti, Fredericka B. (See Liberman, I. Y., SR-24, 17)

________ The Velopharyngeal Mechanism: An Electromyographic Study. A Preliminary Report. SR-25/26 (1971), 117.

________ and H. Hirose. Velopharyngeal Function in Oral/Nasal Articulation and Voicing Gestures. SR-28 (1971), 143.

Blechner, Mark J., R. S. Day, and J. E. Cutting. Processing Two Dimensions of Nonspeech Stimuli: The Auditory-Phonetic Distinction Reconsidered. SR-42/43 (1975), 221.

________ Right-ear Advantage for Musical Stimuli Differing in Rise Time. SR-47 (1976), 63.

________ Left-Ear Advantage for Sounds Characterized by a Rapidly Varying Resonance Frequency. SR-48 (1976), 279.

Borden, Gloria J. Some Effects of Oral Anesthesia Upon Speech: An Electromyographic Investigation. SR-29/30 (1972), 27.

__________, K. S. Harris, and L. Catena. Electromyographic Study of Speech Musculature During Lingual Nerve Block. SR-31/32 (1972), 213.

__________, K. S. Harris, and W. Oliver. Oral Feedback, Part I: Variability of the Effect of Nerve-Block Anesthesia Upon Speech. SR-34 (1973), 159.

__________, K. S. Harris, and L. Catena. Oral Feedback, Part II: An Electromyographic Study of Speech Under Nerve-Block Anesthesia. SR-34 (1973), 167.

__________ and T. Gay. A Combined Cinefluorographic-Electromyographic Study of the Tongue During the Production of /s/: Preliminary Observations. SR-41 (1975), 197.

__________ and T. Gay. Durations of Articulator Movements for /s/-Stop Clusters. SR-44 (1975), 147.

__________ (See Dorman, SR-45/46, 187)

Borst, John M. and H. M. Bailin. Input Equipment for Speech Research Facility. SR-3 (1965), 2.1.

Brady, Susan A. (See Dorman, SR-37/38, 233)

__________ (See Darwin, SR-48, 257)

Brady-Wood, Susan. (See Dorman, SR-33, 93)

__________ and D. Shankweiler. Effects of Amplitude Variation on an Auditory Rivalry Task: Implications Concerning the Mechanism of Perceptual Asymmetries. SR-34 (1973), 119.

Carter, Bonnie. (See Liberman, I. Y., SR-31/32, 145)

Catena, Lorne. (See Borden, SR-31/32, 213)

__________ (See Borden, SR-34, 167)

Chase, R. A. (See MacNeilage, SR-5/6. 5.1)

Collier, Rene. Laryngeal Muscle Activity, Subglottal Air Pressure, and the Control of Pitch in Speech. SR-39/40 (1974), 137.

Cooper, Franklin S. (See Liberman, A. M., SR-1, 2.1)

__________ (See Music, SR-1, 5.1)

__________ (See Liberman, A. M., SR-3, 3.1)

__________ (See Liberman, A. M., SR-4, 1.1)

__________ Computers in Speech Research--Part II. SR-4 (1965), 4.1.

__________ (See Abramson, SR-4, 6.1)

__________ Describing the Speech Process in Motor Command Terms. SR-5/6 (1966), 2.1.

__________ (See Lisker, SR-5/6, 4.1)

__________ (See Liberman, A. M., SR-9, 1.1)

__________ (See Liberman, A. M., SR-10, 1.1)

__________, J. H. Gaitenby, and I. G. Mattingly. High-Performance Reading Machines for the Blind. SR-15/16 (1968), 111.

__________ and I. G. Mattingly. Computer-Controlled PCM System for Investigation of Dichotic Speech Perception. SR-17/18 (1969), 17.

__________, J. H. Gaitenby, I. G. Mattingly, and N. Umeda. Reading Aids for the Blind: A Special Case of Machine-to-Man Communication. SR-17/18 (1969), 117.

__________ (See Liberman, A. M., SR-19/20, 9)

__________ (See Studdert-Kennedy, SR-19/20, 139)

__________ (See Sawashima, SR-19/20, 201)

__________ (See Lisker, SR-21/22, 201)

__________, T. C. Rand, R. S. Music, and I. G. Mattingly. A Voice for the Laboratory Computer. SR-24 (1970), 91.

__________, J. H. Gaitenby, and I. G. Mattingly. Audible Outputs of Reading Machines for the Blind. SR-24 (1970), 97.

__________ How is Language Conveyed by Speech? SR-27 (1971), 3.

__________, J. H. Gaitenby, and I. G. Mattingly. Audible Outputs of Reading Machines for the Blind. SR-28 (1971), 201.

__________, J. H. Gaitenby, I. G. Mattingly, P. W. Nye, and G. N. Sholes. Audible Outputs of Reading Machines for the Blind. SR-29/30 (1972), 91; and SR-35/36 (1973), 117.

__________ (See Nye, SR-29/30, 97)

__________ Machines and Speech. SR-31/32 (1972), 159.

__________ (See Gaitenby, SR-31/32, 177)

__________ (See Gay, SR-33, 221; and SR-35/36, 7)

__________ (See Nye, SR-44, 95)

__________ (See Nye, SR-49, 87)

Copeland, Paul M. (See Day, SR-27, 193)

Crelin, Edmund S. (See Lieberman, SR-21/22, 157)

__________ (See Lieberman, SR-24, 57)

Croog, Joanne, F. (See Turvey, SR-31/32, 67)

Crowder, Robert G. (See Darwin, SR-25/26, 13)

Cullen, John K., Jr. (See Yeni-Komshian, SR-11, 47)

__________ (See Preston, SR-11, 55)

Cutting, James E. (See Day, SR-24, 35)

__________ (See Day, SR-27, 193)

__________ and R. S. Day. Dichotic Fusion Along an Acoustic Continuum. SR-28 (1971), 103.

__________ A Parallel Between Encodedness and the Magnitude of the Right-Ear Effect. SR-29/30 (1972), 61.

__________ (See Day, SR-31/32, 49)

__________ Ear Advantage for Stops and Liquids in Initial and Final Position. SR-31/32 (1972), 57.

__________ A Preliminary Report on Six Fusions in Auditory Research. SR-31/32 (1972), 93.

__________ Phonological Fusion in Synthetic and Natural Speech. SR-33 (1973), 19.

__________ Perception of Speech and Nonspeech, with and without Transitions. SR-33 (1973), 37.

__________ Speech Misperception: Inferences About a Cue for Cluster Perception from a Phonological Fusion Task. SR-33 (1973), 57.

__________ Levels of Processing in Phonological Fusion. (Ph.D. dissertation) SR-34 (1973), 1.

__________ Phonological Fusion of Synthetic Stimuli in Dichotic and Binaural Presentation Modes. SR-34 (1973), 55.

__________ Phonological Fusion of Stimuli Produced by Different Vocal Tracts. SR-34 (1973), 61.

__________ Perception of Speech and Nonspeech, with Speech-Relevant and Speech-Irrelevant Transitions. SR-35/36 (1973), 55.

__________ and P. D. Eimas. Phonetic Feature Analyzers and the Processing of Speech in Infants. SR-37/38 (1974), 45.

__________ (See Kubovy, SR-37/38, 139)

__________ and B. S. Rosner. Categories and Boundaries in Speech and Music. SR-37/38 (1974), 145.

__________ Different Speech-Processing Mechanisms Can Be Reflected in the Results of Discrimination and Dichotic Listening Tasks. SR-37/38 (1974), 159.

__________ (See Dorman, SR-37/38, 217)

__________ Linguistic and Nonlinguistic Stimulus Dimensions Interact in Audition but not in Vision. SR-39/40 (1974), 133.

__________ and J. F. Kavanagh. On the Relationship of Speech to Language. SR-41 (1975), 59.

__________, B. S. Rosner, and C. F. Foard. Rise Time in Nonlinguistic Sounds and Models of Speech Perception. SR-41 (1975), 71.

__________ The Magical Number Two and the Natural Categories of Speech and Music. SR-42/43 (1975), 189.

__________ (See Blechner, SR-42/43, 221)

__________ Predicting Initial Cluster Frequencies by Phonemic Difference. SR-42/43 (1975), 233.

__________ Auditory and Linguistic Processes in Speech Perception: Inferences from Six Fusions in Dichotic Listening. SR-44 (1975), 37.

__________ and M. Dorman. Discrimination of Intensity Differences Carried on Formant Transitions Varying in Extent and Duration. SR-47 (1976), 47.

__________ and B. S. Rosner. Discrimination Functions Predicted from Categories in Speech and Music. SR-47 (1976), 59.

__________ (See Jusczyk, SR-47, 137)

__________ and D. Pisoni. An Information-Processing Approach to Speech Perception. SR-48 (1976), 287.

Darwin, Christopher J. Stimuli for Dichotic Listening Experiments. SR-11 (1967), 61.

__________ Ear Differences in the Recall of Fricatives and Vowels. SR-21/22 (1970), 39.

__________, M. T. Turvey, and R. G. Crowder. An Auditory Analogue of the Sperling Partial Report Procedure: Evidence for Brief Auditory Storage. SR-25/26 (1971), 13.

__________ Stimulus Versus Response Competition and Ear Asymmetry in Dichotic Listening. SR-25/26 (1971), 29.

__________ Dichotic Backward Masking of Complex Sounds. SR-27 (1971), 199.

__________ The Perception of Speech. SR-42/43 (1975), 59.

__________ On the Dynamic Use of Prosody in Speech Perception. SR-42/43 (1975), 103.

__________, P. Howell, and S. A. Brady. Laterality and Localization: A Right Ear Advantage for Speech Heard on the Left. SR-48 (1976), 257.

Day, Ruth S. Temporal Order Judgments in Speech: Are Individuals Language-Bound or Stimulus-Bound? SR-21/22 (1970), 71.

__________ and J. E. Cutting. Perceptual Competition Between Speech and Nonspeech. SR-24 (1970), 35.

__________ Temporal Order Perception of a Reversible Phoneme Cluster. SR-24 (1970), 47.

__________ (See Wood, SR-25/26, 41)

__________ and C. C. Wood. Interactions Between Linguistic and Nonlinguistic Processing. SR-27 (1971), 185.

__________, J. E. Cutting, and P. M. Copeland. Perception of Linguistic and Nonlinguistic Dimensions of Dichotic Stimuli. SR-27 (1971), 193.

__________ and J. C. Bartlett. Separate Speech and Nonspeech Processing in Dichotic Listening? SR-28 (1971), 93.

__________ (See Cutting, SR-28, 103)

__________ and C. C. Wood. Mutual Interference Between Two Linguistic Dimensions of the Same Stimuli. SR-29/30 (1972), 69.

__________ and J. M. Vigorito. A Parallel between Degree of Encodedness and the Ear Advantage: Evidence from a Temporal Order Judgment Task. SR-31/32 (1972), 41.

__________, J. C. Bartlett, and J. E. Cutting. Memory for Dichotic Pairs: Disruption of Ear Report Performance by the Speech-Nonspeech Distinction. SR-31/32 (1972), 49.

__________ Digit-Span Memory in Language-Bound and Stimulus-Bound Subjects. SR-34 (1973), 127.

__________ On Learning "Secret Languages." SR-34 (1973), 141.

__________ (See Blechner, SR-42/43, 221)

Dechovitz, David. (See Lisker, SR-42/43, 257)

DeClerk, Joseph L. (See MacNeilage, SR-12, 9)

Donald, S. Lea. (See Nye, SR-41, 117)

__________ The Effects of Selective Adaptation on Voicing in Thai and English. SR-47 (1976), 129.

__________ (See Gaitenby, SR-48, 327)

Dooley, Patricia M. (See Shelton, SR-12, 79)

Dorman, Michael F. Auditory Evoked Potential Correlates of Speech Sound Discrimination. SR-29/30 (1972), 111.

__________ and R. Hoffman. Short-Term Habituation of the Infant Auditory Evoked Response. SR-29/30 (1972), 121.

__________ Discrimination of Intensity Differences on Formant Transitions In and Out of Syllable Context. SR-33 (1973), 13.

__________, D. Kewley-Port, S. Brady-Wood, and M. T. Turvey. Forward and Backward Masking of Brief Vowels. SR-33 (1973), 93.

__________ and D. S. Geffner. Hemispheric Specialization for Speech Perception in Six-Year-Old Black and White Children from Low and Middle Socioeconomic Classes. SR-34 (1973), 151.

__________ and R. J. Porter, Jr. Hemispheric Lateralization for Speech Perception in Stutterers. SR-37/38 (1974), 117.

__________, J. E. Cutting, and L. J. Raphael. Identification of Vowel Order: Concatenated Versus Formant-Connected Sequences. SR-37/38 (1974), 217.

__________, D. Kewley-Port, S. Brady, and M. T. Turvey. Two Processes in Vowel Recognition: Inferences from Studies of Backward Masking. SR-37/38 (1974), 233.

__________ (See Raphael, SR-37/38, 255)

__________, L. J. Raphael, F. Freeman, and C. Tobin. Vowel and Nasal Durations as Perceptual Cues to Voicing in Word-Final Stop Consonants. SR-37/38 (1974), 263.

__________ (See Geffner, SR-42/43, 241)

__________, L. J. Raphael, A. M. Liberman, and B. Repp. Some Maskinglike Phenomena in Speech Perception. SR-42/43 (1975), 265.

__________ (See Raphael, SR-42/43, 277)

__________, F. J. Freeman, and G. J. Borden. The Effect of Delayed Auditory Feedback on Phonation: An Electromyographic Study. SR-45/46 (1976), 187.

__________, M. Studdert-Kennedy, and L. J. Raphael. Stop Consonant Recognition: Release Bursts and Formant Transitions as Functionally Equivalent, Context-Dependent Cues. SR-47 (1976), 1.

__________ (See Cutting, SR-47, 47)

__________, A. M. Liberman, and L. Raphael. Further Observations on the Role of Silence in the Perception of Stop Consonants. SR-48 (1976), 199.

Edman, Thomas R. (See Strange, SR-45/46, 37)

__________ (See Verbrugge, SR-45/46, 63)

__________ (See Verbrugge, SR-47, 165)

Eimas, Peter D. (See Cutting, SR-37/38, 45)

Elbert, L. (See Music, SR-1, 5.1)

Ellison, G. Campbell. Phonetic Coding of Words in a Taxonomic Classification Task. SR-41 (1975), 95.

Epstein, Robert. A Transistorized Formant-Type Synthesizer. SR-1 (1965), 7.1.

Erickson, Donna and A. S. Abramson. Electromyographic Study of the Tones of Thai. SR-31/32 (1972), 231.

__________, I. G. Mattingly, and M. T. Turvey. Phonetic Coding of Kanji. (Abstract) SR-29/30 (1972), 155.

__________, I. G. Mattingly, and M. T. Turvey. Phonetic Activity in Reading: An Experiment with Kanji. SR-33 (1973), 137.

__________ (See Lisker, SR-42/43, 257)

__________ and J. E. Atkinson. The Function of Strap Muscles in Speech. SR-45/46 (1976), 205.

__________ (See Atkinson, SR-49, 97)

__________, M. Liberman and S. Niimi. The Geniohyoid and the Role of the Strap Muscles. SR-49 (1977), 103.

Fischer, F. William. (See Liberman, I. Y., SR-31/32, 145)

__________ (See Liberman, I. Y., SR-39/40, 105)

Fischer-Jørgensen, Eli and H. Hirose. A Preliminary Electromyographic Study of Labial and Laryngeal Muscles in Danish Stop Consonant Production. SR-39/40 (1974), 231.

__________ and H. Hirose. A Note on Laryngeal Activity in the Danish "stød". SR-39/40 (1974), 255.

Fitch, Hollis. (See Mermelstein, SR-48, 51)

Foard, Christopher F. (See Cutting, SR-41, 71)

__________ (See Jusczyk, SR-47, 137)

Fowler, Carol A. (See Liberman, I. Y., SR-39/40, 105)

__________ A Systems Approach to the Cerebral Hemispheres. SR-44 (1975), 17.

__________, I. Y. Liberman, and D. Shankweiler. On Interpreting the Error Pattern in Beginning Reading. SR-45/46 (1976), 17.

__________ (See Mark, SR-49, 71)

Frazier, Lyn. What Can /w/, /l/, /y/ Tell Us About Categorical Perception? SR-48 (1976), 235.

Freeman, Frances J. and T. Ushijima. Laryngeal Activity Accompanying the Moment of Stuttering: A Preliminary Report of EMG Investigations. SR-37/38 (1974), 109.

__________ (See Raphael, SR-37/38, 255)

__________ (See Dorman, SR-37/38, 263)

__________ and T. Ushijima. The Stuttering Larynx: An EMG, Fiberoptic Study of Laryngeal Activity Accompanying the Moment of Stuttering. SR-41 (1975), 217.

__________ (See Dorman, SR-45/46, 187)

__________ and T. Ushijima. Laryngeal Muscle Activity in Stuttering. SR-45/46 (1976), 211.

Fujimura, O. (See Miyawaki, SR-33, 67)

Gaitenby, Jane H. The Elastic Word. SR-2 (1965), 3.1.

__________ Problems in Machine Conversion of Print to "Speech." SR-5/6 (1966), 7.1.

__________ Rules for Word Stress Analysis for Conversion of Print to Synthetic Speech. SR-12 (1967), 131.

__________ (See Cooper, SR-15/16, 111)

__________ (See Cooper, SR-17/18, 117)

__________ (See Cooper, SR-24, 97)

__________ (See Cooper, SR-28, 201)

__________ (See Cooper, SR-29/30, 91; and SR-35/36, 117)

__________, G. N. Sholes, and G. M. Kuhn. Word and Phrase Stress by Rule for a Reading Machine. SR-29/30 (1972), 105.

__________, G. Sholes, T. Rand, G. Kuhn, P. Nye, and F. Cooper. An Automated Reading Service for the Blind. SR-31/32 (1972), 177.

__________ (See Nye, SR-33, 77)

__________ (See Nye, SR-37/38, 169)

__________ Stress and the Elastic Syllable: An Acoustic Method for Delineating Lexical Stress Patterns in Connected Speech. SR-41 (1975), 137.

__________ and S. L. Donald. Outline of Surname Pronunciation Rules for a Reading Machine. SR-48 (1976), 327.

__________ and P. Mermelstein. Acoustic Correlates of Perceived Prominence in Unknown Utterances. SR-49 (1977), 201.

Garcia, Erica. The Identification and Discrimination of Synthetic Nasals. SR-7/8 (1966), 3.1.

__________ Labelling of Synthetic Nasals (II). SR-9 (1967), 4.1.

__________ Identification and Discrimination of Synthetic Nasals. SR-10 (1967), 8.1.

__________ Discrimination of Three-Formant Nasal-Vowel Syllables (III). SR-12 (1967), 143.

__________ The Perception of Lateral Versus Nasal Syllabic Resonants. SR-13/14 (1968), 117.

__________ The Perception of Liquids, I: /r/ vs. /w/. SR-17/18 (1969), 41.

Gay, Thomas J. A Perceptual Study of English Diphthongs. (Ph.D. dissertation) Supplement to the Status Report, June 1967.

__________ (See Harris, SR-13/14, 137)

__________ Effect of Speaking Rate on Diphthong Formant Movements. SR-15/16 (1968), 97.

__________ Current Status of Electromyography in Speech Research. SR-15/16 (1968), 125.

__________ (See Liberman, SR-17/18, 83)

__________ (See Harris, SR-19/20, 127)

__________ (See Sawashima, SR-19/20, 211)

__________ Effects of Filtering and Vowel Environment on Consonant Perception. SR-21/22 (1970), 133.

__________, H. Hirose, M. Strome, and M. Sawashima. Electromyography of the Intrinsic Laryngeal Muscles During Phonation. SR-25/26 (1971), 97.

__________ (See Hirose, SR-28, 115)

__________ (See Hirose, SR-29/30, 49)

__________ and H. Hirose. Effect of Speaking Rate on Labial Consonant Production: A Combined Electromyographic-High Speed Motion Picture Study. SR-31/32 (1972), 193.

__________, T. Ushijima, H. Hirose, and F. S. Cooper. Effect of Speaking Rate on Labial Consonant-Vowel Articulation. SR-33 (1973), 221; and SR-35/36 (1973), 7.

__________ A Cinefluorographic Study of Vowel Production. SR-39/40 (1974), 171.

__________ and T. Ushijima. Effect of Speaking Rate on Stop Consonant-Vowel Articulation. SR-39/40 (1974), 213.

__________ Jaw Movements During Speech: A Cinefluorographic Investigation. SR-39/40 (1974), 219.

__________ (See Borden, SR-41, 197)

__________ Some Electromyographic Measures of Coarticulation in VCV Utterances. SR-44 (1975), 137.

__________ (See Borden, SR-44, 147)

__________ (See Baer, SR-45/46, 175)

__________ Articulatory Movements in VCV Sequences. SR-49 (1977), 121.

Geffner, Donna S. (See Dorman, SR-34, 151)

__________ and M. F. Dorman. Hemispheric Specialization for Speech Perception in Four-Year-Old Children from Low and Middle Socioeconomic Classes. SR-42/43 (1975), 241.

Gelder, Peter van. (See Rubin, SR-44, 75)

Goff, William R. (See Wood, SR-25/26, 41)

Hadding-Koch, Kerstin and M. Studdert-Kennedy. A Study of Semantic and Psychophysical Test Responses to Controlled Variations in Fundamental Frequency. SR-1 (1965), 4.1.

__________ and A. S. Abramson. Duration Versus Spectrum in Swedish Vowels: Some Perceptual Experiments. SR-2 (1965), 2.1.

__________ and M. Studdert-Kennedy. Further Experimental Studies of Fundamental Frequency Contours. SR-4 (1965), 5.1.

Hadding, Kerstin. (See Studdert-Kennedy, SR-27, 153)

__________ and M. Studdert-Kennedy. Are You Asking Me, Telling Me, or Talking to Yourself? SR-33 (1973), 1.

__________ (See Harris, SR-41, 175)

Haggard, Mark P. A Simple Program for Synthesizing Speech by Rules. SR-9 (1967), 5.1.

__________ Figural After-Effect Phenomena for Speech and Non-Speech Stimuli. SR-9 (1967), 6.1.

__________ and I. G. Mattingly. A Simple Program for Synthesizing British English. SR-12 (1967), 117.

__________ Perception of Semi-Vowels and Laterals. SR-17/18 (1969), 143.

__________ (See Summerfield, SR-49, 1)

Halwes, Terry G. Some Attempts to Make Vowels Behave Like Speech. SR-11 (1967), 57.

__________ (See Mattingly, SR-15/16, 67)

__________ Effects of Dichotic Fusion on the Perception of Speech. (Ph.D. dissertation) Supplement to the Status Report, September 1969.

__________ (See Mattingly, SR-21/22, 99)

Hankins, J. D. (See Nye, SR-29/30, 97)

Harris, Katherine S. (See Liberman, A. M., SR-1, 2.1)

__________, G. F. Lysaught, and M. H. Schvey. Some Aspects of the Production of Oral and Nasal Labial Stops. SR-2 (1965), 1.1.

__________ (See Liberman, A. M., SR-3, 3.1)

__________ (See Huntington, SR-3, 5.1)

__________ (See Liberman, A. M., SR-4, 1.1)

__________ (See Shankweiler, SR-4, 2.1)

__________, D. A. Huntington, and G. N. Sholes. Coarticulation of Some Disyllabic Utterances Measured by Electromyographic Techniques. SR-5/6 (1966), 3.1.

__________ (See Huntington, SR-7/8, 2.1)

__________ (See Shankweiler, SR-9, 2.1)

__________ (See Huntington, SR-10, 6.1)

__________ (See Shelton, SR-12, 79)

__________, T. J. Gay, G. N. Sholes, and P. Lieberman. Some Stress Effects on Electromyographic Measures of Consonant Articulations. SR-13/14 (1968), 137.

__________ (See Lieberman, SR-17/18, 23)

__________ (See Lieberman, SR-17/18, 83)

__________, T. J. Gay, P. Lieberman, and G. N. Sholes. The Function of Muscles in Control of Stress and Intonation. SR-19/20 (1969), 127.

__________ (See Studdert-Kennedy, SR-19/20, 139)

__________ (See Sawashima, SR-19/20, 211)

__________ Physiological Aspects of Articulatory Behavior. SR-23 (1970), 49.

__________ (See Liberman, I. Y., SR-24, 17)

__________ Action of the Extrinsic Musculature in the Control of Tongue Position: Preliminary Report. SR-25/26 (1971), 87.

__________ Vowel Stress and Articulatory Reorganization. SR-28 (1971), 167.

__________ (See Borden, SR-31/32, 213)

__________ (See Borden, SR-34, 159)

__________ (See Borden, SR-34, 167)

__________ (See Bell-Berti, SR-35/36, 1)

__________ Stress and Syllable Duration Change. SR-35/36 (1973), 31.

__________ (See Bell-Berti, SR-37/38, 73)

__________ Mechanisms of Duration Change. SR-39/40 (1974), 193.

__________, H. Hirose, and K. Hadding. Facial Muscle Activity in the Production of Swedish Vowels: An Electromyographic Study. SR-41 (1975), 175.

__________ (See Bell-Berti, SR-42/43, 297)

__________ (See Bell-Berti, SR-45/46, 197)

__________ Physiological Aspects of Speech Production. SR-48 (1976), 21.

Hirose, Hajime. Electromyography of the Articulatory Muscles: Current Instrumentation and Technique. SR-25/26 (1971), 73.

__________ (See Gay, SR-25/26, 97)

__________ An Electromyographic Study of Laryngeal Adjustments During Speech Articulation: A Preliminary Report. SR-25/26 (1971), 107.

__________ and T. J. Gay. The Activity of the Intrinsic Laryngeal Muscles in Voicing Control: An Electromyographic Study. SR-28 (1971), 115.

__________ (See Berti, SR-28, 143)

__________ Laryngeal Adjustments for Vowel Devoicing in Japanese: An Electromyographic Study. SR-28 (1971), 157.

__________ and T. Gay. Laryngeal Control in Vocal Attack: An Electromyographic Study. SR-29/30 (1972), 49.

__________, L. Lisker, and A. S. Abramson. Physiological Aspects of Certain Laryngeal Features in Stop Production. SR-31/32 (1972), 183.

__________ (See Gay, SR-31/32, 193)

__________ (See Bell-Berti, SR-31/32, 207)

__________ (See Bell-Berti, SR-31/32, 223)

__________ (See Gay, SR-33, 221; and SR-35/36, 7)

__________, C. Y. Lee, and T. Ushijima. Laryngeal Control in Korean Stop Production. SR-34 (1973), 191.

__________ (See Bell-Berti, SR-34, 203)

__________ (See Ushijima, SR-37/38, 79)

__________ and T. Ushijima. The Function of the Posterior Cricoarytenoid in Speech Articulation. SR-37/38 (1974), 99.

__________ (See Fischer-Jørgensen, SR-39/40, 231)

__________ (See Fischer-Jørgensen, SR-39/40, 255)

__________ (See Harris, SR-41, 175)

__________ (See Ushijima, SR-41, 207)

Hoffman, Robert. (See Dorman, SR-29/30, 121)

Howell, Peter. (See Darwin, SR-48, 257)

Howie, John M. Experiments on the Perception of Mandarin Vowels and Tones, I. SR-7/8 (1966), 7.1.

__________ Experiments on the Perception of Mandarin Vowels and Tones, II. SR-13/14 (1968), 189.

Huntington, Dorothy A. and K. S. Harris. Some Electromyographic Observations of the Speech of Persons with Impaired Hearing. SR-3 (1965), 5.1.

__________ (See Harris, SR-5/6, 3.1)

__________, K. S. Harris, and G. N. Sholes. An Electromyographic Study of Articulatory Dynamics in Hearing-Impaired and Normal Speakers. SR-7/8 (1966), 2.1.

__________, K. S. Harris, D. P. Shankweiler, and G. N. Sholes. Some Observations on Monosyllable Production by Deaf and Dysarthric Speakers. SR-10 (1967), 6.1.

Ingemann, Frances and P. Mermelstein. Speech Recognition Through Spectrogram Matching. SR-39/40 (1974), 53.

__________ (See Nye, SR-41, 117)

__________ Testing Synthesis-by-Rule with the OVEBORD Program. SR-41 (1975), 127.

Jenkins, J. J. (See Miyawaki, SR-33, 67)

Jusczyk, Peter W., B. S. Rosner, J. E. Cutting, C. F. Foard, and L. B. Smith. Perception of Nonspeech by Infants. SR-47 (1976), 137.

Kavanagh, James F. (See Mattingly, SR-29/30, 83)

__________ (See Cutting, SR-41, 59)

Kewley-Port, Diane and M. S. Preston. Early Apical Stop Production: A Voice Onset Time Analysis. SR-29/30 (1972), 125.

__________ (See Turvey, SR-31/32, 131)

__________ (See Dorman, SR-33, 93)

__________ Computer Processing of EMG Signals at Haskins Laboratories. SR-33 (1973), 173.

__________ An Experimental Evaluation of the EMG Data Processing System: Time Constant Choice for Digital Integration. SR-37/38 (1974), 65.

__________ (See Dorman, SR-37/38, 233)

__________ Word Recall in Aphasia. SR-39/40 (1974), 123.

Kirstein, Emily F. Perception of Second-Formant Transitions in Non-Speech Patterns. SR-7/8 (1966), 91.

__________ and D. P. Shankweiler. Selective Listening for Dichotically Presented Stop Consonants and Vowels. SR-17/18 (1969), 133.

__________ Selective Listening for Temporally Staggered Dichotic CV Syllables. SR-21/22 (1970), 63.

__________ Perception of Dichotically Presented Steady-State Vowels as a Function of Interaural Delay. SR-24 (1970), 31.

__________ Temporal Factors in Perception of Dichotically Presented Stop Consonants and Vowels. (Ph.D. dissertation) SR-24 (1970), Part II, 1.

__________ The Lag Effect in Dichotic Speech Perception. SR-35/36 (1973), 81.

Klatt, Dennis H. (See Lieberman, SR-15/16, 157)

__________ (See Lieberman, SR-24, 57)

Knudson, Ronald. (See Lieberman, SR-13/14, 153)

Kubovy, Michael, J. E. Cutting, and R. McI. McGuire. Binaural Subjective Tones and Melodies Without Monaural Familiarity Cues. SR-37/38 (1974), 139.

Kuhn, Gary M. The Phi Coefficient as an Index of Ear Differences in Dichotic Listening. SR-29/30 (1972), 75.

__________ (See Gaitenby, SR-29/30, 105)

__________ (See Gaitenby, SR-31/32, 177)

__________ and R. McI. McGuire. Results of a VCV Spectrogram-Reading Experiment. SR-39/40 (1974), 67.

__________ On the Front Cavity Resonance, and Its Possible Role in Speech Perception. SR-41 (1975), 105.

Lee, Charles Y. (See Hirose, SR-34, 191)

Leidner, David R. An Electromyographic Study of the American English Liquids. SR-33 (1973), 195.

Levitt, Andrea G. (See Nearey, SR-39/40, 81)

Liberman, Alvin M., F. S. Cooper, K. S. Harris, P. F. MacNeilage and M. Studdert-Kennedy. Some Observations on a Model for Speech Perception. SR-1 (1965), 2.1.

__________, M. Studdert-Kennedy, K. S. Harris, and F. S. Cooper. A Reply to "Identification and Discrimination Functions for a Visual Continuum and Their Relation to the Motor Theory of Speech Perception," by Cross, Lane, and Sheppard. SR-3 (1965), 3.1.

__________, F. S. Cooper, M. Studdert-Kennedy, K. S. Harris, and D. P. Shankweiler. Some Observations on the Efficiency of Speech Sounds. SR-4 (1965), 1.1.

__________, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. Perception of the Speech Code. SR-9 (1967), 1.1.

__________, F. S. Cooper, and M. Studdert-Kennedy. Why Are Spectrograms Hard to Read? SR-10 (1967), 1.1.

__________ (See Stevens, SR-13/14, 47)

__________ (See Mattingly, SR-15/16, 9)

__________ Some Characteristics of Perception in the Speech Mode. SR-15/16 (1968), 39.

__________ (See Mattingly, SR-15/16, 67)

__________ and F. S. Cooper. In Search of the Acoustic Cues. SR-19/20 (1969), 9.

__________ The Grammars of Speech and Language. SR-19/20 (1969), 79.

__________ (See Studdert-Kennedy, SR-19/20, 139)

__________ (See Mattingly, SR-21/22, 99)

__________, I. G. Mattingly, and M. T. Turvey. Language Codes and Memory Codes. SR-27 (1971), 59.

__________ The Specialization of the Language Hemisphere. SR-31/32 (1972), 1.

__________ (See Miyawaki, SR-33, 67)

__________ State-of-the-Art Report on Language Processing. SR-35/36 (1973), 107.

__________ The Role of Speech in Language: Introduction to the Conference. SR-37/38 (1974), 1.

__________ (See Liberman, I. Y., SR-39/40, 105)

__________ (See Lisker, SR-42/43, 257)

__________ (See Dorman, SR-42/43, 265)

__________ (See Raphael, SR-42/43, 277)

__________ How Abstract Must a Motor Theory of Speech Perception Be? SR-44 (1975), 1.

__________ Comments on the Session: Perception and Production of Speech II; Conference on Origins and Evolution of Language and Speech. SR-45/46, 29.

__________ and D. B. Pisoni. Evidence for a Special Speech-Perceiving Subsystem in the Human. SR-48 (1976), 183.

__________ (See Dorman, SR-48, 199)

Liberman, Isabelle Y., D. P. Shankweiler, C. Orlando, K. S. Harris, and F. B. Berti. Letter Confusions and Reversals of Sequence in the

Beginning Reader: Implications for Orton's Theory of Developmental Dyslexia. SR-24 (1970), 17. (See also Erratum, SR-27, 211)

__________ Basic Research in Speech and Lateralization of Language: Some Implications for Reading Disability. SR-25/26 (1971), 51.

__________ (See Shankweiler, SR-27, 35)

__________, D. Shankweiler, B. Carter, and F. W. Fischer. Reading and the Awarenesss of Linguistic Segments. SR-31/32 (1972), 145.

__________ Segmentation of the Spoken Word and Reading Acquisition. SR-33 (1973), 157.

__________, D. Shankweiler, A. M. Liberman, C. Fowler, and F. W. Fischer. Phonetic Segmentation and Recoding in the Beginning Reader. SR-39/40 (1974), 105.

__________ (See Shankweiler, SR-45/46, 1)

__________ (See Fowler, SR-45/46, 17)

__________ and D. Shankweiler. Speech, the Alphabet and Teaching to Read. SR-48 (1976), 75.

__________ (See Mark, SR-49, 71)

Liberman, Mark. (See Erickson, SR-49, 103)

Lieberman, Philip. Primate Vocalizations and Human Linguistic Ability. SR-13/14 (1968), 65.

__________ (See Harris, SR-13/14, 137)

__________, R. Knudson, and J. Mead. Determination of the Rate of Change of Fundamental Frequency with Respect to Subglottal Air Pressure During Sustained Phonation. SR-13/14 (1968), 153.

__________ On the Physical Bases of Phonologic Features. SR-15/16 (1968), 137.

__________, D. H. Klatt, and W. A. Wilson. Vocal Tract Limitations on the Vowel Repertoires of Rhesus Monkey and Other Nonhuman Primates. SR-15/16 (1968), 157.

__________, K. S. Harris, P. Wolff, and L. R. Russell. Newborn Infant Cry and Nonhuman Primate Vocalizations. SR-17/18 (1969), 23.

__________, M. Sawashima, K. S. Harris, and T. J. Gay. An Articulatory Implementation of the Breath-Group and Prominence: Crico-Thyroid Muscular Activity in Intonation. SR-17/18 (1969), 83.

__________ Towards a Unified Phonetic Theory. SR-17/18 (1969), 157.

__________ (See Harris, SR-19/20, 127)

__________ and E. S. Crelin. On the Speech of Neanderthal Man. SR-21/22 (1970), 157.

__________ A Study of Prosodic Features. SR-23 (1970), 179.

__________, E. S. Crelin, and D. H. Klatt. Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and the Chimpanzee. SR-24 (1970), 57.

__________ On the Evolution of Human Language. SR-27 (1971), 113.

__________ The Evolution of Human Speech Anatomy. SR-28 (1971), 205.

__________ Linguistic and Paralinguistic Interchange. SR-33 (1973), 167.

__________ On the Evolution of Language: A Unified View. SR-33 (1973), 229.

__________ The Evolution of Speech and Language. SR-37/38 (1974), 25.

Lindblom, Björn E. F. and M. Studdert-Kennedy. On the Role of Formant Transitions in Vowel Recognition. SR-9 (1967), 7.1.

Lisker, Leigh and A. S. Abramson. Voice Onset Time in the Production and Perception of English Stops. SR-1 (1965), 3.1.

__________ (See Abramson, SR-3, 1.1)

__________ Supraglottal Air Pressure in the Production of English Stops. SR-4 (1965), 3.1.

__________ (See Abramson, SR-4, 6.1)

__________ and A. S. Abramson. Some Effects of Context on Voice Onset Time in English Stops. SR-5/6 (1966), 1.1.

__________, A. S. Abramson, F. S. Cooper, and M. H. Schvey. Transillumination of the Larynx in Running Speech. SR-5/6 (1966), 4.1.

__________ Measuring Stop Closure Duration from Intraoral Pressure Records. SR-7/8 (1966), 5.1.

__________ Pattern to "Holes" in the Pattern? SR-10 (1967), 5.1.

__________ and A. S. Abramson. The Voicing Dimension: Some Experiments in Comparative Phonetics. SR-11 (1967), 9.

__________ (See Abramson, SR-11, 17)

__________ (See Abramson, SR-11, 23)

__________ (See Abramson, SR-13/14, 49)

__________ and A. S. Abramson. Distinctive Features and Laryngeal Control. SR-15/16 (1968), 81.

__________ Stop Duration and Voicing in English. SR-19/20 (1969), 27.

__________ (See Sawashima, SR-19/20, 201)

__________, M. Sawashima, A. S. Abramson, and F. S. Cooper. Cinegraphic Observations of the Larynx During Voiced and Voiceless Stops. SR-21/22 (1970), 201.

__________ On Time and Timing in Speech. SR-23 (1970), 151.

__________ On Learning a New Contrast. SR-24 (1970), 1.

__________ and A. S. Abramson. Distinctive Features and Laryngeal Control. SR-27 (1971), 133.

__________ and A. S. Abramson. Glottal Modes in Consonant Distinctions. SR-27 (1971), 175.

__________ (See Abramson, SR-27, 179)

__________ (See Abramson, SR-29/30, 15)

__________ (See Hirose, SR-31/32, 183)

__________ On "Explaining" Vowel Duration Variation. SR-37/38 (1974), 225.

__________ Is it VOT or a First-Formant Transition Detector? SR-41 (1975), 153.

__________, A. M. Liberman, D. Erickson, and D. Dechovitz. On Pushing the Voice-Onset-Time (VOT) Boundary About. SR-42/43 (1975), 257.

__________ In (Qualified) Defense of VOT. SR-44 (1975), 109.

__________ Phonetic Aspects of Time and Timing. SR-47 (1976), 113.

__________ Stop Voicing Production: Natural Outputs and Synthesized Inputs. SR-47 (1976), 155.

Lorge, Beatrice. A Study of the Relationship Between Production and Perception of Initial and Intervocalic /t/ and /d/ in Individual English-Speaking Adults. SR-9 (1967), 3.1.

__________ Some Rules for Synthesis of General American English. SR-13/14 (1968), 37.

__________ and A. M. Liberman. The Speech Code and the Physiology of Language. SR-15/16 (1968), 9.

__________, A. M. Liberman, A. K. Syrdal, and T. Halwes. Discrimination of F_2 Transitions in Speech Context and in Isolation. SR-15/16 (1968), 67.

__________ (See Cooper, SR-15/16, 111)

__________ Programming for the Glace-Holmes Synthesizer. SR-17/18 (1969), 9.

__________ (See Cooper, SR-17/18, 17)

__________ (See Cooper, SR-17/18, 117)

__________ Synthesis by Rule as a Tool for Phonological Research. SR-19/20 (1969), 61.

__________, A. M. Liberman, A. K. Syrdal, and T. Halwes. Discrimination in Speech and Nonspeech Modes. SR-21/22 (1970), 99.

__________ Speech Synthesis for Phonetic and Phonological Models. SR-23 (1970), 117.

__________ (See Cooper, SR-24, 91)

__________ (See Cooper, SR-24, 97)

__________ Reading, the Linguistic Process, and Linguistic Awareness. SR-27 (1971), 23.

__________ (See Liberman, A. M., SR-27, 59)

__________ Speech Cues and Sign Stimuli. SR-27 (1971), 89.

__________ (See Cooper, SR-28, 201)

__________ and J. F. Kavanagh. The Relationships Between Speech and Reading. SR-29/30 (1972), 83.

__________ (See Cooper, SR-29/30, 91; and SR-35/36, 117)

__________ (See Nye, SR-29/30, 97)

__________ (See Erickson, SR-29/30, 155)

__________ (See Erickson, SR-33, 137)

__________ Phonetic Prerequisites for First-Language Acquisition. SR-34 (1973), 65.

__________ The Human Aspect of Speech. SR-37/38 (1974), 5.

__________ Syllable Synthesis. SR-49 (1977), 111.

May, Janet. Vocal Tract Normalization for /s/ and /š/. SR-48 (1976), 67.

Mead, Jere. (See Lieberman, SR-13/14, 153)

Mermelstein, Paul. A Phonetic-Context Controlled Strategy for Segmentation and Phonetic Labeling of Speech. SR-37/38 (1974), 191.

__________ (See Ingemann, SR-39/40, 53)

__________ Automatic Segmentation of Speech into Syllabic Units. SR-42/43 (1975), 247.

__________ On Detecting Nasals in Continuous Speech. SR-44 (1975), 83.

__________ (See Nye, SR-44, 95)

__________ Distance Measures for Speech Recognition - Psychological and Instrumental. SR-47 (1976), 91.

__________ and H. Fitch. Difference Limens for Formant Frequencies for Steady-State and Consonant-Bounded Vowels. SR-48 (1976), 51.

__________ (See Nye, SR-49, 87)

__________ (See Gaitenby, SR-49, 201)

Michaels, Claire Farley and M. T. Turvey. Hemiretinae and Nonmonotonic Masking Functions with Overlapping Stimuli. SR-31/32 (1972), 127.

__________ An Examination of Hemispheric Asymmetry in the Visual Processing of Linguistic Terms. (Ph.D dissertation) Abstract. SR-35/36 (1973), 121.

Miyawaki, Kuniko, A. M. Liberman, O. Fujimura, W. Strange, and J. J. Jenkins. Cross-Language Study of the Perception of the F_3 Cue for [r] versus [l] in Speech- and Nonspeech-Like Patterns. SR-33 (1973), 67.

Montlick, Terry. (See Nye, SR-44, 95)

Morse, Philip Allen. The Discrimination of Speech and Nonspeech Stimuli in Early Infancy. (Ph.D. dissertation) Abstract. SR-29/30 (1972), 151.

Music, Richard S., L. Elbert, N. Reilly, and F. S. Cooper. Automating the Collection and Reduction of Electromyographic Data on Speech Production, I and II. SR-1 (1965), 5.1.

__________ (See Cooper, SR-24, 91)

Nearey, Terrance M. (See Nye, SR-37/38, 123)

__________ and A. G. Levitt. Evidence for Spectral Fusion in Dichotic Release from Upward Spread of Masking. SR-39/40 (1974), 81.

Niimi, Seiji. (See Baer, SR-45/46, 175)

__________ (See Erickson, SR-49, 103)

Nye, Patrick W. (See Cooper, SR-29/30, 91; and SR-35/36, 117)

__________, J. D. Hankins, T. Rand, I. G. Mattingly, and F. S. Cooper. Field Evaluation of an Automated Reading System for the Blind. SR-29/30 (1972), 97.

__________ (See Gaitenby, SR-31/32, 177)

__________ and J. H. Gaitenby. Consonant Intelligibility in Synthetic Speech and in a Natural Speech Control (Modified Rhyme Test Results). SR-33 (1973), 77.

__________, T. M. Nearey, and T. C. Rand. Dichotic Release from Masking: Further Results from Studies with Synthetic Speech Stimuli. SR-37/38 (1974), 123.

__________ and J. H. Gaitenby. The Intelligibility of Synthetic Monosyllabic Words in Short, Syntactically Normal Sentences. SR-37/38 (1974), 169.

__________, F. Ingemann, and L. Donald. Synthetic Speech Comprehension: A Comparison of Listener Performances with and Preferences Among Different Speech Forms. SR-41 (1975), 117.

__________, L. J. Reiss, F. S. Cooper, R. M. McGuire, P. Mermelstein, and T. Montlick. A Digital Pattern Playback for the Analysis and Manipulation of Speech Signals. SR-44 (1975), 95.

__________, F. S. Cooper, and P. Mermelstein. Interactive Experiments with a Digital Pattern Playback. SR-49 (1977), 87.

Öhman, Sven E. G. (See Stevens, SR-13/14, 47)

Oliver, William. (See Borden, SR-34, 159)

Orlando, Charles. (See Liberman, I. Y., SR-24, 17)

Phillips, Juliet. (See Preston, SR-13/14, 185)

Pisoni, David B. (See Studdert-Kennedy, SR-25/26, 1)

__________ On the Nature of Categorical Perception of Speech Sounds. (Ph.D. dissertation) Abstract: SR-27 (1971), 209. Text: Supplement to Status Report, November 1971.

__________ (See Turvey, SR-31/32, 67)

__________ Perceptual Processing Time for Consonants and Vowels. SR-31/32 (1972), 83.

__________ and Jeffrey Tash. Reaction Times to Comparisons Within and Across Phonetic Categories: Evidence for Auditory and Phonetic Levels of Processing. SR-34 (1973), 77.

__________ The Role of Auditory Short-Term Memory in Vowel Perception. SR-34 (1973), 89.

__________ and J. R. Sawusch. On the Identification of Place and Voicing Features in Synthetic Stop Consonants. SR-35/36 (1973), 65.

__________ (See Liberman, SR-48, 183)

__________ (See Cutting, SR-48, 287)

Port, Diane K. (See Preston, SR-13/14, 181)

__________ (See Preston, SR-19/20, 189)

__________ The EMG Data System. SR-25/26 (1971), 67.

Port, Diane Kewley. (See Kewley-Port)

Porter, Robert J. The Effect of Temporal Overlap on the Perception of Dichotically and Monotically Presented CV Syllables. SR-25/26 (1971), 35.

__________ The Effect of Delayed Channel on the Perception of Dichotically Presented Speech and Nonspeech Sounds. (Ph.D. dissertation) Abstract. SR-29/30 (1972), 153.

__________ (See Dorman, SR-37/38, 117)

Preston, Malcolm S. A Study of Voicing in the Stops Found in the Pre-Linguistic Babblings of Infants. SR-7/8 (1966), 6.1.

__________, G. Yeni-Komshian, and R. E. Stark. A Study of Voicing in Initial Stops Found in the Pre-Linguistic Vocalizations of Infants from Different Language Environments. SR-10 (1967), 7.1.

__________ (See Winterkorn, SR-11, 41)

__________ (See Yeni-Komshian, SR-11, 47)

__________ and G. Yeni-Komshian. Studies on the Development of Stop-Consonants in Children. SR-11 (1967), 49.

__________ and J. K. Cullen. Computer Synthesis of Infant Vocalizations. SR-11 (1967), 55.

__________, G. Yeni-Komshian, and P. Benson. A Dichotic ABX Procedure for Use in Laterality Studies. SR-13/14 (1968), 179.

__________, G. Yeni-Komshian, R. E. Stark, and D. K. Port. Developmental Studies of Voicing in Stops. SR-13/14 (1968), 181.

__________, J. Phillips, and G. Yeni-Komshian. Procedures for Examining the Perceptual Capabilities for Processing Speech Sounds in Children Under Two Years of Age. SR-13/14 (1968), 185.

__________ and D. K. Port. Further Results on the Development of Voicing in Stop-Consonants in Young Children. SR-19/20 (1969), 139.

__________ (See Kewley-Port, SR-29/30, 125)

Prindle, Sandra Sears. (See Turvey, SR-47, 29)

Rand, Timothy C. (See Cooper, SR-24, 91)

__________ Vocal Tract Size Normalization in the Perception of Stop Consonants. SR-25/26 (1971), 141.

__________ (See Nye, SR-29/30, 97)

__________ (See Gaitenby, SR-31/32, 177)

__________ Dichotic Release from Masking for Speech. SR-33 (1973), 47.

__________ (See Nye, SR-37/38, 123)

Raphael, Lawrence J. An Electromyographic Investigation of the Tense-Lax Feature in Some English Vowels. SR-25/26 (1971), 131.

__________ Vowel Duration as a Cue to the Perceptual Separation of Cognate Sounds in American English. (Ph.D. dissertation) Abstract: SR-25/26 (1971), 147. Text: Supplement to the Status Report, August 1971.

__________ An Electromyographic Investigation of the Features of Tension in Some American English Vowels. SR-28 (1971), 179.

__________ and F. Bell-Berti. The Role of the Extrinsic and Intrinsic Tongue Muscles in Differentiating the English Tense-Lax Vowel Pairs. SR-33 (1973), 203.

__________ (See Dorman, SR-37/38, 217)

__________, M. F. Dorman, C. Tobin, and F. Freeman. Vowel and Nasal Durations in Vowel-Nasal-Consonant Sequences in American English: Spectrographic Studies. SR-37/38 (1974), 255.

__________ (See Dorman, SR-37/38, 263)

__________ The Physiological Control of Durational Differences between Vowels Preceding Voiced and Voiceless Consonants in English. SR-39/40 (1974), 199.

__________ (See Dorman, SR-42/43, 265)

__________, M. F. Dorman, and A. M. Liberman. The Perception of Vowel Duration in VC and CVC Syllables. SR-42/43 (1975), 277.

__________ (See Dorman, SR-47, 1)

__________ (See Dorman, SR-48, 199)

Reilly, Norman. (See Music, SR-1, 5.1)

Reiss, L. J. (See Nye, SR-44, 95)

Repp, Bruno H. "Coperception": A Preliminary Study. SR-42/43 (1975), 147.

__________ Dichotic "Masking" of Voice Onset Time. SR-42/43 (1975), 159.

__________ (See Dorman, SR-42/43, 265)

__________ Identification of Dichotic Fusions. SR-45/46 (1976), 95.

__________ Discrimination of Dichotic Fusions. SR-45/46 (1976), 123.

__________ Coperception: Two Further Preliminary Studies. SR-45/46 (1976), 141.

__________ "Posner's Paradigm" and Categorical Perception: A Negative Study. SR-45/46 (1976), 153.

__________ Dichotic Competition of Speech Sounds: The Role of Acoustic Stimulus Structure. SR-47 (1976), 71.

__________ Perception of Implosive Transitions in VCV Utterances. SR-48 (1976), 209.

__________ Perceptual Integration and Selective Attention in Speech Perception: Further Experiments on Intervocalic Stop Consonants. SR-49 (1977), 37.

__________ Measuring Laterality Effects in Dichotic Listening. SR-49 (1977), 149.

__________ A Simple Model of Response Selection in the Dichotic Two-Response Paradigm. SR-49 (1977), 187.

Rootes, Thomas P. and P. F. MacNeilage. Coordinated Studies of a Syndrome of Impairment in Somesthetic Perception and Motor Function. SR-1 (1965), 8.1.

__________ (See MacNeilage, SR-5/6, 5.1)

Rosner, Burton S. (See Cutting, SR-37/38, 145)

__________ (See Cutting, SR-41, 71)

__________ (See Cutting, SR-47, 59)

__________ (See Jusczyk, SR-47, 137)

Rubin, Philip, M. T. Turvey, and Peter van Gelder. Initial Phonemes Are Detected Faster in Spoken Words than in Spoken Nonwords. SR-44 (1975), 75.

Russell, Lorraine H. (See Lieberman, SR-17/18, 23)

Sawashima, Masayuki. (See Lieberman, SR-17/18, 83)

__________, A. S. Abramson, F. S. Cooper, and L. Lisker. Observing Laryngeal Adjustments During Running Speech by Use of a Fiberoptics System.

SR-19/20 (1969), 201.

__________, T. J. Gay, and K. S. Harris. Laryngeal Muscle Activity During Vocal Pitch and Intensity Changes. SR-19/20 (1969), 211.

__________ Glottal Adjustments for English Obstruents. SR-21/22 (1970), 187.

__________ (See Lisker, SR-21/22, 201)

__________ Laryngeal Research in Experimental Phonetics. SR-23 (1970), 69.

__________ (See Gay, SR-25/26, 97)

Sawusch, James R. (See Pisoni, SR-35/36, 65)

Schvey, Malcolm H. (See Harris, SR-2, 1.1)

__________ (See Lisker, SR-5/6, 4.1)

Schulman, Susan. (See Studdert-Kennedy, SR-21/22, 89)

Shankweiler, Donald P. Laterality Effects in Perception of Speech and Other Sounds. SR-3 (1965), 6.1.

__________ (See Liberman, A. M., SR-4, 1.1)

__________ and K. S. Harris. An Experimental Approach to the Problem of Articulation in Aphasia. SR-4 (1965), 2.1.

__________ and M. Studdert-Kennedy. Identification of Consonants and Vowels Presented to Left and Right Ears. SR-5/6 (1966), 6.1.

__________ Hemispheric Specialization and the Perception of Speech. SR-7/8 (1966), 1.1.

__________ (See Liberman, A. M., SR-9, 1.1)

__________, K. S. Harris, and M. L. Taylor. Electromyographic Studies of Articulation in Aphasia. SR-9 (1967), 2.1.

__________ and M. Studdert-Kennedy. An Analysis of Perceptual Confusions in Identification of Dichotically Presented CVC Syllables. SR-10 (1967), 4.1.

__________ (See Huntington, SR-10, 6.1)

__________ An Analysis of Laterality Effects in Speech Perception. SR-13/14 (1968), 9.

__________ (See Kirstein, SR-17/18, 133)

__________ (See Studdert-Kennedy, SR-21/22, 1)

__________ (See Studdert-Kennedy, SR-21/22, 89)

__________ (See Liberman, I. Y., SR-24, 17)

__________ (See Studdert-Kennedy, SR-25/26, 1)

__________ and I. Y. Liberman. Misreading: A Search for Causes. SR-27 (1971), 35.

__________ (See Studdert-Kennedy, SR-31/32, 23)

__________ (See Liberman, I. Y., SR-31/32, 145)

__________ (See Brady-Wood, SR-34, 119)

__________ (See Verbrugge, SR-37/38, 199)

__________ (See Strange, SR-37/38, 209)

__________ (See Liberman, SR-39/40, 105)

__________, W. Strange, and R. Verbrugge. Speech and the Problem of Perceptual Constancy. SR-42/43 (1975), 117.

__________, W. Strange, and R. Verbrugge. On Accounting for the Poor Recognition of Isolated Vowels. SR-42/43 (1975), 285.

__________ and I. Y. Liberman. Exploring the Relations between Reading and Speech. SR-45/46 (1976), 1.

__________ (See Fowler, SR-45/46, 17)

__________ (See Strange, SR-45/46, 37)

__________ (See Verbrugge, SR-45/46, 63)

__________ (See Verbrugge, SR-47, 165)

__________ (See Liberman, SR-48, 75)

__________ (See Mark, SR-49, 71)

Shaw, Robert. (See Turvey, SR-48, 1)

Shelton, Ralph L., K. S. Harris, G. N. Sholes, and P. M. Dooley. Study of Non-Speech Voluntary Palate Movements by Scaling and Electromyographic Techniques. SR-12 (1967), 79.

Sholes, George N. Automating the Collection and Reduction of Electromyographic Data on Speech Production, III: Measuring and Averaging the Data. SR-1 (1965), 6.1.

__________ (See Harris, SR-5/6, 3.1)

__________ (See Huntington, SR-7/8, 2.1)

__________ (See Huntington, SR-10, 6.1)

__________ (See Shelton, SR-12, 79)

__________ (See Harris, SR-13/14, 137)

__________ (See Harris, SR-19/20, 127)

__________ (See Cooper, SR-29/30, 91; and SR-35/36, 117)

__________ (See Gaitenby, SR-29/30, 105)

__________ (See Gaitenby, SR-31/32, 177)

__________ Weak Syllables in a Primitive Reading-Machine Algorithm. SR-45/46 (1976), 163.

__________ Building an -S Detection and Removal Algorithm. SR-48 (1976), 339.

Singer, Andrew J. The Design and Implication of a Narrative Data Reduction System. Supplement to the Status Report, October 1966.

Smith, Linda B. (See Jusczyk, SR-47, 137)

Stark, Rachel E. (See Preston, SR-10, 7.1)

__________ (See Preston, SR-13/14, 181)

Stevens, Kenneth N., A. M. Liberman, M. Studdert-Kennedy, and S. E. G. Öhman. Cross-Language Study of Vowel Perception (A). SR-13/14 (1968), 47.

Strange, Winifred. (See Miyawaki, SR-33, 67)

__________ (See Verbrugge, SR-37/38, 199)

__________, R. Verbrugge, and D. Shankweiler. Consonant Environment Specifies Vowel Identity. SR-37/38 (1974), 209.

__________ (See Shankweiler, SR-42/43, 117)

__________ (See Shankweiler, SR-42/43, 285)

__________, R. R. Verbrugge, D. P. Shankweiler, and T. R. Edman. Consonant Environment Specifies Vowel Identity. SR-45/46 (1976), 37.

__________ (See Verbrugge, SR-45/46, 63)

__________ (See Verbrugge, SR-47, 165)

Strome, Marshall. (See Gay, SR-25/26, 97)

Studdert-Kennedy, Michael. (See Liberman, A. M., SR-1, 2.1)

__________ (See Hadding-Koch, SR-1, 4.1)

__________ (See Liberman, A. M., SR-3, 3.1)

__________ (See Liberman, A. M., SR-4, 1.1)

__________ (See Hadding-Koch, SR-4, 5.1)

__________ (See Shankweiler, SR-5/6, 6.1)

__________ (See Liberman, A. M., SR-9, 1.1)

__________ (See Liberman, A. M., SR-10, 1.1)

__________ (See Shankweiler, SR-10, 4.1)

__________ (See Stevens, SR-13/14, 47)

__________, A. M. Liberman, K. S. Harris, and F. S. Cooper. A Motor Theory of Speech Perception: A Reply to Lane's Critical Review. SR-19/20 (1969), 139.

__________ and D. P. Shankweiler. Hemispheric Specialization for Speech Perception. SR-21/22 (1970), 1.

__________, D. P. Shankweiler, and S. Schulman. Opposed Effects of a Delayed Channel on Perception of Dichotically and Monotically Presented CV Syllables. SR-21/22 (1970), 89.

__________ The Perception of Speech. SR-23 (1970), 15.

__________, D. P. Shankweiler, and D. B. Pisoni. Auditory and Phonetic Processes in Speech Perception: Evidence from a Dichotic Study. SR-25/26 (1971), 1.

__________ and K. Hadding. Auditory and Linguistic Processes in the Perception of Intonation Contours. SR-27 (1971), 153.

__________ and D. Shankweiler. A Continuum of Cerebral Dominance for Speech Perception? SR-31/32 (1972), 23.

__________ A Right-Ear Advantage in Choice Reaction Time to Monaurally Presented Vowels: A Pilot Study. SR-31/32 (1972), 75.

__________ (See Hadding, SR-33, 1)

__________ From Continuous Signal to Discrete Message: Syllable to Phoneme. SR-37/38 (1974), 13.

__________ Speech Perception. SR-39/40 (1974), 1.

__________ Two Questions in Dichotic Listening. SR-41 (1975), 51.

__________ (See Dorman, SR-47, 1)

__________ Universals in Phonetic Structure and Their Role in Linguistic Communication. SR-48 (1976), 43.

Sullivan, Edie V. and M. T. Turvey. On the Short-Term Retention of Serial, Tactile Stimuli. SR-33 (1973), 123.

Summerfield, Quentin and M. Haggard. On the Dissociation of Spectral and Temporal Cues to the Voicing Distinction in Initial Stop Consonants. SR-49 (1977), 1.

Syrdal, Ann K. (See Mattingly, SR-15/16, 67)

__________ (See Mattingly, SR-21/22, 99)

Tash, Jeffrey. (See Pisoni, SR-34, 77)

Taylor, Martha L. (See Shankweiler, SR-9, 2.1)

Tobin, Charles. (See Raphael, SR-37/38, 255)

__________ (See Dorman, SR-37/38, 263)

Turvey, Michael T. (See Darwin, SR-25/26, 13)

__________ (See Liberman, A. M., SR-27, 59)

__________ On Peripheral and Central Processes in Vision: Inferences from an Information-Processing Analysis of Masking with Patterned Stimuli. SR-28 (1971), 1.

__________ Some Aspects of Selective Readout from Iconic Storage. SR-29/30 (1972), 1.

__________ (See Erickson, SR-29/30, 155)

__________, D. Pisoni, and J. F. Croog. A Right-Ear Advantage in the Retention of Words Presented Monaurally. SR-31/32 (1972), 67.

__________ Constructive Theory, Perceptual Systems, and Tacit Knowledge. SR-31/32 (1972), 109.

__________ (See Michaels, SR-31/32, 127)

__________, C. F. Michaels, and D. Kewley-Port. Visual Storage or Visual Masking?: An Analysis of the "Retroactive Contour Enhancement" Effect. SR-31/32 (1972), 131.

__________ (See Dorman, SR-33, 93)

__________ and R. A. Weeks. Effects of Proactive Interference and Rehearsal on the Primary and Secondary Components of Short-Term Retention. SR-33 (1973), 101.

__________ (See Sullivan, SR-33, 123)

__________ (See Erickson, SR-33, 137)

__________ A Note on the Relation Between Action and Perception. SR-34 (1973), 71.

__________ (See Dorman, SR-37/38, 233)

__________ Preliminaries to a Theory of Action with Reference to Vision. SR-41 (1975), 1.

__________ Perspectives in Vision: Conception or Perception? SR-42/43 (1975), 1.

__________ (See Rubin, SR-44, 75)

__________ and S. S. Prindle. Modes of Perceiving: Abstracts, Comments and Notes. SR-47 (1976), 29.

__________, R. Shaw, and W. Mace. Issues in the Theory of Action. SR-48 (1976), 1.

__________ Visual Processing and Short-Term Memory. SR-48 (1976), 97.

___________ Contrasting Orientations to the Theory of Visual Information-Processing. SR-48 (1976), 153.

Umeda, Noriko. (See Cooper, SR-17/18, 117)

Ushijima, Tatsujiro. (See Gay, SR-33, 221; and SR-35/36, 7)

___________ (See Hirose, SR-34, 191)

___________ and H. Hirose. Electromyographic Study of the Velum During Speech. SR-37/38 (1974), 79.

___________ (See Hirose, SR-37/38, 99)

___________ (See Freeman, SR-37/38, 109)

___________ (See Gay, SR-39/40, 213)

___________ and H. Hirose. Velar Movement and Its Motor Command. SR-41 (1975), 207.

___________ (See Freeman, SR-41, 217)

___________ (See Freeman, SR-45/46, 211)

Verbrugge, Robert, W. Strange, and D. Shankweiler. What Information Enables a Listener to Map a Talker's Vowel Space? SR-37/38 (1974), 199.

___________ (See Strange, SR-37/38, 209)

___________ (See Shankweiler, SR-42/43, 117)

___________ (See Shankweiler, SR-42/43, 285)

___________ (See Strange, SR-45/46, 37)

___________, W. Strange, D. P. Shankweiler, and T. R. Edman. What Information Enables a Listener to Map a Talker's Vowel Space? SR-45/46 (1976), 63.

___________, D. Shankweiler, W. Strange, and T. R. Edman. Shifts in Vowel Perception as a Function of Speaking Rate. SR-47 (1976), 165.

Vigorito, James M. (See Day, SR-31/32, 41)

Vinegrad, Michael D. The Scaling of Speech Sounds. SR-7/8 (1966), 8.1.

__________ The Scaling of Synthetic Speech. SR-11 (1967), 37.

__________ A Direct Magnitude Scaling Method to Investigate Categorical Versus Continuous Modes of Speech Perception. SR-21/22 (1970), 147.

Weeks, Robert A. A Speech Perception Paradox?: The Right-Ear Advantage and the Lag Effect. SR-33 (1973), 29.

__________ (See Turvey, SR-33, 101)

Wilson, William A. (See Lieberman, SR-15/16, 157)

Winterkorn, J. M. S., P. F. MacNeilage, and M. S. Preston. Perception of Voiced and Voiceless Stop Consonants in Three-Year-Old Children. SR-11 (1967), 41.

Wolff, Peter. (See Lieberman, SR-17/18, 23)

Wood, Charles C., W. R. Goff, and R. S. Day. Auditory Evoked Potentials During Speech Perception. SR-25/26 (1971), 41.

__________ (See Day, SR-27, 185)

__________ (See Day, SR-29/30, 69)

__________ Levels of Processing in Speech Perception: Neurophysiological and Information-Processing Analyses. (Ph.D. dissertation) Abstract--SR-34 (1973), 215; Text--Supplement bound with SR-35/36 (1973), 123.

__________ Parallel Processing of Auditory and Phonetic Information in Speech Perception. SR-35/36 (1973), 39.

Yeni-Komshian, Grace. (See Preston, SR-10, 7.1)

__________, M. S. Preston, and J. K. Cullen. A Brief Report on a Study of Imitation of Synthetic Speech Stimuli by Children. SR-11 (1967), 47.

__________ (See Preston, SR-11, 49)

__________ (See Preston, SR-13/14, 179)

__________ (See Preston, SR-13/14, 181)

__________ (See Preston, SR-13/14, 185)

TITLE INDEX

Haskins Laboratories Status Report on Speech Research

SR-1 (1965) - SR-49 (1977)

On Accounting for the Poor Recognition of Isolated Vowels. D. Shankweiler, W. Strange, and R. Verbrugge. SR-42/43 (1975), 285.

Acoustic Correlates of Perceived Prominence in Unknown Utterances. J. H. Gaitenby and P. Mermelstein. SR-49 (1977), 201.

Action of the Extrinsic Musculature in the Control of Tongue Position: Preliminary Report. K. S. Harris. SR-25/26 (1971), 87.

The Activity of the Intrinsic Laryngeal Muscles in Voicing Control: An Electromyographic Study. H. Hirose and T. J. Gay. SR-28 (1971), 115.

An Analysis of Laterality Effects in Speech Perception. D. P. Shankweiler. SR-13/14 (1968), 9.

An Analysis of Perceptual Confusions in Identification of Dichotically Presented CVC Syllables. D. P. Shankweiler. SR-10 (1967), 4.1.

Are You Asking Me, Telling Me, or Talking to Yourself? K. Hadding and M. Studdert-Kennedy. SR-33 (1973), 1.

An Articulatory Implementation of the *Breath-Group* and *Prominence*: Cricothyroid Muscular Activity in Intonation. P. Lieberman, M. Sawashima, K. S. Harris, and T. J. Gay. SR-17/18 (1968), 83.

Articulatory Movements in VCV Sequences. T. Gay. SR-49 (1977), 121.

Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency. J. E. Atkinson. (Ph.D. dissertation) Abstract. SR-34 (1973), 211.

Audible Outputs of Reading Machines for the Blind. F. S. Cooper, J. H. Gaitenby, and I. G. Mattingly. SR-24 (1970), 97.

Audible Outputs of Reading Machines for the Blind. F. S. Cooper, J. H. Gaitenby, and I. G. Mattingly. SR-28 (1971), 201.

Audible Outputs of Reading Machines for the Blind. F. S. Cooper, J. H. Gaitenby, I. G. Mattingly, P. W. Nye, and G. N. Sholes. SR-29/30 (1972), 91; and SR-35/36 (1973), 117.

An Auditory Analogue of the Sperling Partial Report Procedure: Evidence for Brief Auditory Storage. C. J. Darwin, M. T. Turvey, and R. G. Crowder. SR-25/26 (1971), 13.

Auditory Evoked Potential Correlates of Speech Sound Discrimination. M. F. Dorman. SR-29/30 (1972), 111.

Auditory Evoked Potentials During Speech Perception. C. C. Wood, W. R. Goff, and R. S. Day. SR-25/26 (1971), 41.

Auditory and Linguistic Processes in the Perception of Intonation Contours. M. Studdert-Kennedy and K. Hadding. SR-27 (1971), 153.

Auditory and Linguistic Processes in Speech Perception: Inferences from Six Fusions in Dichotic Listening. J. E. Cutting. SR-44 (1975), 37.

Auditory and Phonetic Processes in Speech Perception: Evidence from a Dichotic Study. M. Studdert-Kennedy, D. P. Shankweiler, and D. B. Pisoni. SR-25/26 (1971), 1.

An Automated Reading Service for the Blind. J. Gaitenby, G. Sholes, T. Rand, G. Kuhn, P. Nye, and F. Cooper. SR-31/32 (1972), 177.

Automatic Segmentation of Speech into Syllabic Units. P. Mermelstein. SR-42/43 (1975), 247.

Automating the Collection and Reduction of Electromyographic Data on Speech Production I and II. R. S. Music, L. Elbert, N. Reilly, and F. S. Cooper. SR-1 (1965), 5.1.

Automating the Collection and Reduction of Electromyographic Data on Speech Production, III: Measuring and Averaging the Data. G. N. Sholes. SR-1 (1965), 6.1.

Basic Research in Speech and Lateralization of Language: Some Implications for Reading Disability. I. Y. Liberman. SR-25/26 (1971), 51.

Binaural Subjective Tones and Melodies Without Monaural Familiarity Cues. M. Kubovy, J. E. Cutting, and R. McI. McGuire. SR-37/38 (1974), 139.

Book Review: Mary R. Haas, Thai-English Student's Dictionary. A. S. Abramson. SR-3 (1965), 4.1.

A Brief Report on a Study of Imitation of Synthetic Speech Stimuli by Children. G. Yeni-Komshian, M. S. Preston, and J. K. Cullen. SR-11 (1967), 47.

Building an -S Detection and Removal Algorithm. G. Sholes. SR-48 (1976), 339.

Categorical Perception Along an Oral-Nasal Continuum. R. Mandler. SR-47 (1976), 147.

Categories and Boundaries in Speech and Music. J. E. Cutting and B. S. Rosner. SR-37/38 (1974), 145.

A Cinefluorographic Study of Vowel Production. T. Gay. SR-39/40 (1974), 171.

Cinegraphic Observations of the Larynx During Voiced and Voiceless Stops. L. Lisker, M. Sawashima, A. S. Abramson, and F. S. Cooper. SR-21/22 (1970), 201.

Coarticulation of Some Disyllabic Utterances Measured by Electromyographic Techniques. K. S. Harris, D. A. Huntington, and G. N. Sholes. SR-5/6 (1966), 3.1.

The Coarticulation of Tones: An Acoustic Study of Thai. A. S. Abramson. SR-44 (1975), 119.

A Combined Cinefluorographic-Electromyographic Study of the Tongue During the Production of /s/: Preliminary Observations. G. Borden and T. Gay. SR-41 (1975), 197.

Comments on the Session: Perception and Production of Speech II; Conference on Origins and Evolution of Language and Speech. A. M. Liberman. SR-45/46 (1976), 29.

Computer-Controlled PCM System for Investigation of Dichotic Speech Perception. F. S. Cooper and I. G. Mattingly. SR-17/18 (1969), 17.

Computer Processing of EMG Signals at Haskins Laboratories. D. Kewley-Port. SR-33 (1973), 173.

Computer Synthesis of Infant Vocalizations. M. S. Preston and J. K. Cullen. SR-11 (1967), 55.

Computers in Speech Research -- Part II. F. S. Cooper. SR-4 (1965), 4.1.

Consonant Environment Specifies Vowel Identity. W. Strange, R. Verbrugge, and D. Shankweiler. SR-37/38 (1974), 209.

Consonant Environment Specifies Vowel Identity. W. Strange, R. R. Verbrugge, D. P. Shankweiler, and T. R. Edman. SR-45/46 (1976), 37.

Consonant Intelligibility in Synthetic Speech and in a Natural Speech Control (Modified Rhyme Test Results). P. W. Nye and J. H. Gaitenby. SR-33 (1973), 77.

Constructive Theory, Perceptual Systems, and Tacit Knowledge. M. T. Turvey. SR-31/32 (1972), 109.

A Continuum of Cerebral Dominance for Speech Perception? M. Studdert-Kennedy and D. Shankweiler. SR-31/32 (1972), 23.

Contrasting Orientations to the Theory of Visual Information Processing. M. T. Turvey. SR-48 (1976), 153.

Control of Fundamental Frequency, Intensity, and Register of Phonation. T. Baer, T. Gay, and S. Niimi. SR-45/46 (1976), 175.

Coordinated Studies of a Syndrome of Impairment of Somesthetic Perception and Motor Function. T. P. Rootes and P. F. MacNeilage. SR-1 (1965), 8.1.

"Coperception": A Preliminary Study. B. H. Repp. SR-42/43 (1975), 147.

Coperception: Two Further Preliminary Studies. B. H. Repp. SR-45/46 (1976), 141.

Cross-Language Study of the Perception of the F3 Cue for [r] versus [l] in Speech- and Nonspeech-Like Patterns. K. Miyawaki, A. M. Liberman, O. Fujimura, W. Strange, and J. J. Jenkins. SR-33 (1973), 67.

Cross-Language Study of Vowel Perception (A). K. N. Stevens, A. M. Liberman, M. Studdert-Kennedy, and S. E. G. Öhman. SR-13/14 (1968), 47.

Current Status of Electromyography in Speech Research. T. J. Gay. SR-15/16 (1968), 125.

Describing the Speech Process in Motor Command Terms. F. S. Cooper. SR-5/6 (1966), 2.1.

The Design and Implementation of a Narrative Data Reduction System. A. J. Singer. Supplement to the Status Report, October 1966.

On Detecting Nasals in Continuous Speech. P. Mermelstein. SR-44 (1975), 83.

Determination of the Rate of Change of Fundamental Frequency with Respect to Subglottal Air Pressure During Sustained Phonation. P. Lieberman. SR-13/14 (1968), 153.

Developmental Studies of Voicing in Stops. M. S. Preston, G. Yeni-Komshian, R. E. Stark, and D. K. Port. SR-13/14 (1968), 181.

A Dichotic ABX Procedure for Use in Laterality Studies. M. S. Preston, G. Yeni-Komshian, and P. Benson. SR-13/14 (1968), 179.

Dichotic Backward Masking of Complex Sounds. C. J. Darwin. SR-27 (1971), 199.

Dichotic Competition of Speech Sounds: The Role of Acoustic Stimulus Structure. B. H. Repp. SR-47 (1976), 71.

Dichotic Fusion Along an Acoustic Continuum. J. E. Cutting and R. S. Day. SR-28 (1971), 103.

Dichotic "Masking" of Voice Onset Time. B. H. Repp. SR-42/43 (1975), 159.

Dichotic Release from Masking for Speech. T. C. Rand. SR-33 (1973), 47.

Dichotic Release from Masking: Further Results from Studies with Synthetic Speech Stimuli. P. W. Nye, T. M. Nearey, and T. C. Rand. SR-37/38 (1974), 123.

Difference Limens for Formant Frequencies for Steady-State and Consonant-Bounded Vowels. P. Mermelstein and H. Fitch. SR-48 (1976), 51.

Different Speech-Processing Mechanisms Can Be Reflected in the Results of Discrimination and Dichotic Listening Tasks. J. E. Cutting. SR-37/38 (1974), 159.

Digit-Span Memory in Language-Bound and Stimulus-Bound Subjects. R. S. Day. SR-34 (1973), 127.

A Digital Pattern Playback for the Analysis and Manipulation of Speech Signals. P. W. Nye, L. J. Reiss, F. S. Cooper, R. M. McGuire, P. Mermelstein, and T. Montlick. SR-44 (1975), 95.

A Direct Magnitude Scaling Method to Investigate Categorical Versus Continuous Modes of Speech Perception. M. D. Vinegrad. SR-21/22 (1970), 147.

Discriminability Along the Voicing Continuum: Cross-Language Tests. A. S. Abramson and L. Lisker. SR-11 (1967), 17.

Discrimination of Dichotic Fusions. B. H. Repp. SR-45/46 (1976), 123.

Discrimination of F_2 Transitions in Speech Context and in Isolation. I. G. Mattingly, A. M. Liberman, A. K. Syrdal, and T. Halwes. SR-15/16 (1968), 67.

Discrimination Functions Predicted from Categories in Speech and Music. J. E. Cutting and B. S. Rosner. SR-47 (1976), 59.

Discrimination of Intensity Differences Carried on Formant Transitions Varying in Extent and Duration. J. E. Cutting and M. Dorman. SR-47 (1976), 47.

Discrimination of Intensity Differences on Formant Transitions In and Out of Syllable Context. M. F. Dorman. SR-33 (1973), 13.

Discrimination in Speech and Nonspeech Modes. I. G. Mattingly, A. M. Liberman, A. K. Syrdal, and T. Halwes. SR-21/22 (1970), 99.

The Discrimination of Speech and Nonspeech Stimuli in Early Infancy. P. A. Morse. (Ph.D. dissertation) Abstract. SR-29/30 (1972), 151.

Discrimination of Three-Formant Nasal-Vowel Syllables (III). E. Garcia. SR-12 (1967), 143.

On the Dissociation of Spectral and Temporal Cues to the Voicing Distinction in Initial Stop Consonants. Q. Summerfield and M. Haggard. SR-49 (1977), 1.

Distance Measures for Speech Recognition - Psychological and Instrumental. P. Mermelstein. SR-47 (1976), 91.

Distinctive Features and Laryngeal Control. L. Lisker and A. S. Abramson. SR-15/16 (1968), 81.

Distinctive Features and Laryngeal Control. L. Lisker and A. S. Abramson. SR-27 (1971), 133.

Duration Versus Spectrum in Swedish Vowels: Some Perceptual Experiments. K. Hadding-Koch and A. S. Abramson. SR-2 (1965), 2.1.

Durations of Articulator Movements for /s/-Stop Clusters. G. J. Borden and T. Gay. SR-44 (1975), 147.

On the Dynamic Use of Prosody in Speech Perception. C. J. Darwin. SR-42/43 (1975), 103.

Ear Advantage for Stops and Liquids in Initial and Final Position. J. E. Cutting. SR-31/32 (1972), 57.

Ear Differences in the Recall of Fricatives and Vowels. C. J. Darwin. SR-21/22 (1970), 39.

Early Apical Stop Production: A Voice Onset Time Analysis. D. K. Port and M. S. Preston. SR-29/30 (1972), 125.

The Effect of Delayed Auditory Feedback on Phonation: An Electromyographic Study. M. F. Dorman, F. J. Freeman, and G. J. Borden. SR-45/46 (1976), 187.

The Effect of Delayed Channel on the Perception of Dichotically Presented Speech and Nonspeech Sounds. R. J. Porter, Jr. (Ph.D. dissertation) Abstract. SR-29/30 (1972), 153.

Effect of Speaking Rate on Diphthong Formant Movements. T. J. Gay. SR-15/16 (1969), 97.

Effect of Speaking Rate on Labial Consonant Production: A Combined Electromyographic-High Speed Motion Picture Study. T. Gay and H. Hirose. SR-31/32 (1972), 193.

Effect of Speaking Rate on Labial Consonant-Vowel Articulation. T. Gay, T. Ushijima, H. Hirose, and F. S. Cooper. SR-33 (1973), 221; and SR-35/36 (1973), 7.

Effect of Speaking Rate on Stop Consonant-Vowel Articulation. T. Gay and T. Ushijima. SR-39/40 (1974), 213.

The Effect of Temporal Overlap on the Perception of Dichotically and Monotically Presented CV Syllables. R. J. Porter. SR-25/26 (1971), 35.

Effects of Amplitude Variation on an Auditory Rivalry Task: Implications Concerning the Mechanisms of Perceptual Asymmetries. S. Brady-Wood and D. Shankweiler. SR-34 (1973), 119.

Effects of Dichotic Fusion on the Perception of Speech. T. Halwes. (Ph.D. dissertation). Supplement to the Status Report, September 1969.

Effects of Filtering and Vowel Environment on Consonant Perception. T. J. Gay. SR-21/22 (1970), 133.

Effects of Proactive Interference and Rehearsal on the Primary and Secondary Components of Short-Term Retention. M. T. Turvey and R. A. Weeks. SR-33 (1973), 101.

The Effects of Selective Adaptation on Voicing in Thai and English. S. L. Donald. SR-47 (1976), 129.

The Elastic Word. J. H. Gaitenby. SR-2 (1965), 3.1.

An Electromyographic Investigation of the Feature of Tension in Some American English Vowels. L. J. Raphael. SR-28 (1971), 179.

An Electromyographic Investigation of the Tense-Lax Feature in Some English Vowels. L. J. Raphael. SR-25/26 (1971), 131.

Electromyographic Studies of Articulation in Aphasia. D. Shankweiler, K. S. Harris, and M. L. Taylor. SR-9 (1967), 2.1.

An Electromyographic Study of the American English Liquids. D. R. Leidner. SR-33 (1973), 195.

An Electromyographic Study of Articulatory Dynamics in Hearing-Impaired and Normal Speakers. D. A. Huntington, K. S. Harris, and G. N. Sholes. SR-7/8 (1966), 2.1.

An Electromyographic Study of Laryngeal Adjustments During Speech Articulation: A Preliminary Report. H. Hirose. SR-25/26 (1971), 107.

Electromyographic Study of Speech Musculature During Lingual Nerve Block. G. J. Borden, K. S. Harris, and L. Catena. SR-31/32 (1972), 213.

Electromyographic Study of the Tones of Thai. D. Erickson and A. S. Abramson. SR-31/32 (1972), 231.

Electromyographic Study of the Velum During Speech. T. Ushijima and H. Hirose. SR-37/38 (1974), 79.

Electromyography of the Articulatory Muscles: Current Instrumentation and Technique. H. Hirose. SR-25/26 (1971), 73.

Electromyography of the Intrinsic Laryngeal Muscles During Phonation. T. J. Gay, H. Hirose, M. Strome, and M. Sawashima. SR-25/26 (1971), 97.

The EMG Data System. D. K. Port. SR-25/26 (1971), 67.

Evidence for a Special Speech-Perceiving Subsystem in the Human. A. M. Liberman and D. B. Pisoni. SR-48 (1976), 183.

Evidence for Spectral Fusion in Dichotic Release from Upward Spread of Masking. T. M. Nearey and A. G. Levitt. SR-39/40 (1974), 81.

On the Evolution of Human Language. P. Lieberman. SR-27 (1971), 113.

On the Evolution of Language: A Unified View. P. Lieberman. SR-33 (1973), 229.

The Evolution of Human Speech Anatomy. P. Lieberman. SR-28 (1971), 205.

The Evolution of Speech and Language. P. Lieberman. SR-37/38 (1974), 25.

An Examination of Hemispheric Asymmetry in the Visual Processing of Linguistic Items. C. F. Michaels. (Ph.D. dissertation) Abstract. SR-35/36 (1973), 121.

The Experimental Approach to the Problem of Articulation in Aphasia. D. P. Shankweiler and K. S. Harris. SR-4 (1965), 2.1.

An Experimental Evaluation of the EMG Data Processing System: Time Constant Choice for Digital Integration. D. Kewley-Port. SR-37/38 (1974), 65.

Experimental Methods for Speech Synthesis by Rule. I. G. Mattingly. SR-12 (1967), 103.

Experiments on the Perception of Mandarin Vowels and Tones, I. J. M. Howie. SR-7/8 (1966), 7.1.

Experiments on the Perception of Mandarin Vowels and Tones, II. J. M. Howie. SR-13/14 (1968), 189.

On "Explaining" Vowel Duration Variation. L. Lisker. SR-37/38 (1974), 225.

Exploring the Relations between Reading and Speech. D. Shankweiler and I. Y. Liberman. SR-45/46 (1976), 1.

Facial Muscle Activity in the Production of Swedish Vowels: An Electromyographic Study. K. S. Harris, H. Hirose, and K. Hadding. SR-41 (1975), 175.

Field Evaluation of an Automated Reading System for the Blind. P. W. Nye, J. D. Hankins, T. Rand, I. G. Mattingly, and F. S. Cooper. SR-29/30 (1972), 97.

Figural After-Effect Phenomena for Speech and Non-Speech Stimuli. M. P. Haggard. SR-9 (1967), 6.1.

Forward and Backward Masking of Brief Vowels. M. Dorman, D. Kewley-Port, S. Brady-Wood, and M. T. Turvey. SR-33 (1973), 93.

From Continuous Signal to Discrete Message: Syllable to Phoneme. M. Studdert-Kennedy. SR-37/38 (1974), 13.

On the Front Cavity Resonance, and Its Possible Role in Speech Perception. G. M. Kuhn. SR-41 (1975), 105.

The Function of Muscles in Control of Stress and Intonation. K. S. Harris, T. J. Gay, P. Lieberman, and G. N. Sholes. SR-19/20 (1966), 127.

The Function of the Posterior Cricoarytenoid in Speech Articulation. H. Hirose and T. Ushijima. SR-37/38 (1974), 99.

The Function of Strap Muscles in Speech. D. Erickson and J. E. Atkinson. SR-45/46 (1976), 205.

The Function of Strap Muscles in Speech: Pitch Lowering or Jaw Opening? J. E. Atkinson and D. Erickson. SR-49 (1977), 97.

Further Experimental Studies of Fundamental Frequency Contours. K. Hadding-Koch and M. Studdert-Kennedy. SR-4 (1965), 5.1.

Further Observations on the Role of Silence in the Perception of Stop Consonants. M. F. Dorman, L. J. Raphael and A. M. Liberman. SR-48 (1976), 199.

Further Results on the Development of Voicing in Stop Consonants in Young Children. M. S. Preston and D. K. Port. SR-19/20 (1969), 189.

The Geniohyoid and the Role of the Strap Muscles. D. Erickson, M. Liberman, and S. Niimi. SR-49 (1977), 103.

Glottal Adjustments for English Obstruents. M. Sawashima. SR-21/22 (1970), 187.

Glottal Modes in Consonant Distinctions. L. Lisker and A. S. Abramson. SR-27 (1971), 175.

The Grammars of Speech and Language. A. M. Liberman. SR-19/20 (1969), 79.

Hemiretinae and Nonmonotonic Masking Functions with Overlapping Stimuli. C. F. Michaels and M. T. Turvey. SR-31/32 (1972), 127.

Hemispheric Lateralization for Speech Perception in Stutterers. M. F. Dorman and R. J. Porter, Jr. SR-37/38 (1974), 117.

Hemispheric Specialization and the Perception of Speech. D. P. Shankweiler. SR-7/8 (1966), 1.1.

Hemispheric Specialization for Speech Perception. M. Studdert-Kennedy and D. P. Shankweiler. SR-21/22 (1970), 1.

Hemispheric Specialization for Speech Perception in Four-Year-Old Children from Low and Middle Socioeconomic Classes. D. S. Geffner and M. F. Dorman. SR-42/43 (1975), 241.

Hemispheric Specialization for Speech Perception in Six-Year-Old Black and White Children from Low and Middle Socioeconomic Classes. M. F. Dorman and D. S. Geffner. SR-34 (1973), 151.

High-Performance Reading Machines for the Blind. F. S. Cooper, J. H. Gaitenby, and I. G. Mattingly. SR-15/16 (1968), 111.

How Abstract Must a Motor Theory of Speech Perception Be? A. M. Liberman. SR-44 (1975), 1.

How is Language Conveyed by Speech? F. S. Cooper. SR-27 (1971), 3

The Human Aspect of Speech. I. G. Mattingly. SR-37/38 (1974), 5.

Identification of Consonants and Vowels Presented to Left and Right Ears. D. P. Shankweiler and M. Studdert-Kennedy. SR-5/6 (1966), 6.1.

Identification of Dichotic Fusions. B. H. Repp. SR-45/46 (1976), 95.

The Identification and Discrimination of Synthetic Nasals. E. Garcia. SR-7/8 (1966), 3.1.

Identification and Discrimination of Synthetic Nasals. E. Garcia. SR-10 (1967), 8.1.

On the Identification of Place and Voicing Features in Synthetic Stop Consonants. D. B. Pisoni and J. R. Sawusch. SR-35/36 (1973), 65.

Identification of Vowel Order: Concatenated Versus Formant-Connected Sequences. M. F. Dorman, J. E. Cutting, and L. J. Raphael. SR-37/38 (1974), 217.

An Information-Processing Approach to Speech Perception. J. E. Cutting and D. B. Pisoni. SR-48 (1976), 287.

Initial Phonemes Are Detected Faster in Spoken Words than in Spoken Nonwords. P. Rubin, M. T. Turvey, and Peter van Gelder. SR-44 (1975), 75.

Input Equipment for Speech Research Facility. J. Borst and H. M. Bailin. SR-3 (1965), 2.1.

The Intelligibility of Synthetic Monosyllabic Words in Short, Syntactically Normal Sentences. P. W. Nye and J. H. Gaitenby. SR-37/38 (1974), 169.

Interactions Between Linguistic and Nonlinguistic Processing. R. S. Day and C. C. Wood. SR-27 (1971), 85.

Interactive Experiments with a Digital Pattern Playback. P. W. Nye, F. S. Cooper, and P. Mermelstein. SR-49 (1977), 87.

On Interpreting the Error Pattern in Beginning Reading. C. A. Fowler, I. Y. Liberman, and D. Shankweiler. SR-45/46 (1976), 17.

Is it VOT or a First-Formant Transition Detector? L. Lisker. SR-41 (1975), 153.

Issues in the Theory of Action. M. T. Turvey, R. E. Shaw, and W. Mace. SR-48 (1976), 1.

Jaw Movements During Speech: A Cinefluorographic Investigation. T. Gay. SR-39/40 (1974), 219.

Labelling of Synthetic Nasals (II). E. Garcia. SR-9 (1967), 4.1.

The Lag Effect in Dichotic Speech Perception. E. F. Kirstein. SR-35/36 (1973), 81.

Language Codes and Memory Codes. A. M. Liberman. SR-27 (1971), 59.

Laryngeal Activity Accompanying the Moment of Stuttering: A Preliminary Report of EMG Investigations. F. J. Freeman and T. Ushijima. SR-37/38 (1974), 109.

Laryngeal Activity in Stop Consonants. A. S. Abramson, L. Lisker, and F. S. Cooper. SR-4 (1965), 6.1.

Laryngeal Adjustments for Vowel Devoicing in Japanese: An Electromyographic Study. H. Hirose. SR-28 (1971), 157.

Laryngeal Behavior, the Speech Signal, and Phonological Simplicity. A. S. Abramson and L. Lisker. SR-11 (1967), 23.

Laryngeal Control in Korean Stop Production. H. Hirose, C. Y. Lee, and T. Ushijima. SR-34 (1973), 191.

Laryngeal Control in Vocal Attack: An Electromyographic Study. H. Hirose and T. Gay. SR-29/30 (1972), 49.

Laryngeal Muscle Activity During Vocal Pitch and Intensity Changes. M. Sawashima, T. J. Gay, and K. S. Harris. SR-19/20 (1969), 211.

Laryngeal Muscle Activity in Stuttering. F. J. Freeman and T. Ushijima. SR-45/46 (1976), 211.

Laryngeal Muscle Activity, Subglottal Air Pressure, and the Control of Pitch in Speech. R. Collier. SR-39/40 (1974), 137.

Laryngeal Research in Experimental Phonetics. M. Sawashima. SR-23 (1970), 69.

Laryngeal Timing in Consonant Distinctions. A. Abramson. SR-47 (1976), 105.

Laterality Effects in Perception of Speech and Other Sounds. D. P. Shankweiler. SR-3 (1965), 6.1.

Laterality and Localization: A Right Ear Advantage for Speech Heard on the Left. C. J. Darwin, P. Howell, and S. A. Brady. SR-48 (1976), 257.

On Learning a New Contrast. L. Lisker. SR-24 (1970), 1.

On Learning "Secret Languages." R. S. Day. SR-34 (1973), 141.

Left-Ear Advantage for Sounds Characterized by a Rapidly Varying Resonance Frequency. M. J. Blechner. SR-48 (1976), 279.

Letter Confusions and Reversals of Sequence in the Beginning Reader: Implications for Orton's Theory of Developmental Dyslexia. I. Y. Liberman, D. P. Shankweiler, C. Orlando, K. S. Harris, and F. B. Berti. SR-24 (1970), 17. (See also Erratum, SR-27, 211)

Levels of Processing in Phonological Fusion. J. E. Cutting. (Ph.D. dissertation) SR-34 (1973), 1.

Levels of Processing in Speech Perception: Neurophysiological and Information-Processing Analyses. C. C. Wood. (Ph.D. dissertation) Abstract--SR-34 (1973), 215; Text--Supplement bound with SR-35/36 (1973), 123.

Linguistic and Nonlinguistic Stimulus Dimensions Interact in Audition but not in Vision. J. E. Cutting. SR-39/40 (1974), 133.

Linguistic and Paralinguistic Interchange. P. Lieberman. SR-33 (1973), 167.

Machines and Speech. F. S. Cooper. SR-31/32 (1972), 159.

The Magical Number Two and the Natural Categories of Speech and Music. J. E. Cutting. SR-42/43 (1975), 189.

On Matching the Speech Signal to Mechanisms of Production. A. S. Abramson. SR-10 (1967), 2.1.

Measuring Laterality Effects in Dichotic Listening. B. H. Repp. SR-49 (1977), 149.

Measuring Stop Closure Duration from Intraoral Pressure Records. L. Lisker. SR-7/8 (1966), 5.1.

Mechanisms of Duration Change. K. S. Harris. SR-39/40 (1974), 193.

Memory for Dichotic Pairs: Disruption of Ear Report Performance by the Speech-Nonspeech Distinction. R. S. Day, J. C. Bartlett, and J. E. Cutting. SR-31/32 (1972), 49.

Misreading: A Search for Causes. D. P. Shankweiler and I. Y. Liberman. SR-27 (1971), 35.

Modes of Perceiving: Abstracts, Comments and Notes. M. T. Turvey and S. S. Prindle. SR-47 (1976), 29.

More on the Motor Organization of Speech Gestures. F. Bell-Berti and K. S. Harris. SR-37/38 (1974), 73.

On the Motor Control of Coarticulation in CVC Monosyllables. P. F. MacNeilage and J. L. DeClerk. SR-12 (1967), 9.

The Motor Organization of Some Speech Gestures. F. Bell-Berti and K. S. Harris. SR-35/36 (1973), 1.

A Motor Theory of Speech Perception: A Reply to Lane's Critical Review. M. Studdert-Kennedy, A. M. Liberman, K. S. Harris, and F. S. Cooper. SR-19/20 (1969), 139.

Mutual Interference Between Two Linguistic Dimensions of the Same Stimuli. R. S. Day and C. C. Wood. SR-29/30 (1972), 69.

On the Nature of Categorical Perception of Speech Sounds. D. B. Pisoni. (Ph.D. dissertation) Abstract: SR-27 (1971), 209. Text: Supplement to the Status Report, November 1971.

Newborn Infant Cry and Nonhuman Primate Vocalizations. P. Lieberman, K. S. Harris, P. Wolff, and L. R. Russell. SR-17/18 (1969), 23.

A Note on Laryngeal Activity in the Danish "stød". E. Fischer-Jørgensen and H. Hirose. SR-39/40 (1974), 255.

A Note on the Relation Between Action and Perception. M. T. Turvey. SR-34 (1973), 71.

Observing Laryngeal Adjustments During Running Speech by Use of a Fiberoptics System. M. Sawashima, A. S. Abramson, F. S. Cooper, and L. Lisker. SR-19/20 (1969), 201.

Opposed Effects of a Delayed Channel on Perception of Dichotically and Monotonically Presented CV Syllables. M. Studdert-Kennedy, D. P. Shankweiler, and S. Schulman. SR-21/22 (1970), 89.

Oral Feedback, Part I: Variability of the Effect of Nerve-Block Anesthesia Upon Speech. G. J. Borden, K. S. Harris, and W. Oliver. SR-34 (1973), 159.

Oral Feedback, Part II: An Electromyographic Study of Speech Under Nerve-Block Anesthesia. G. J. Borden, K. S. Harris, and L. Catena. SR-34 (1973), 167.

Outline of Surname Pronunciation Rules for a Reading Machine. J. H. Gaitenby and S. L. Donald. SR-48 (1976), 327.

A Parallel Between Degree of Encodedness and the Ear Advantage: Evidence from a Temporal Order Judgment Task. R. S. Day and J. M. Vigorito. SR-31/32 (1972), 41.

A Parallel Between Encodedness and the Magnitude of the Right-Ear Effect. J. E. Cutting. SR-29/30 (1972), 61.

Parallel Processing of Auditory and Phonetic Information in Speech Perception. C. C. Wood. SR-35/36 (1973), 39.

Pattern to "Holes" in the Pattern? L. Lisker. SR-10 (1967), 5.1.

Patterns of Palatoglossus Activity and Their Implications for Speech Organization. F. Bell-Berti and H. Hirose. SR-34 (1973), 203.

Perception of Dichotically Presented Steady-State Vowels as a Function of Interaural Delay. E. F. Kirstein. SR-24 (1970), 31.

Perception of Implosive Transitions in VCV Utterances. B. H. Repp. SR-48 (1976), 209.

The Perception of Lateral Versus Nasal Syllabic Resonants. E. Garcia. SR-13/14 (1968), 117.

Perception of Linguistic and Nonlinguistic Dimensions of Dichotic Stimuli. R. S. Day, J. E. Cutting, and P. M. Copeland. SR-27 (1971), 193.

The Perception of Liquids, I: /r/ vs. /w/. E. Garcia. SR-17/18 (1969), 41.

Perception of Nonspeech by Infants. P. W. Jusczyk, B. S. Rosner, J. E. Cutting, C. F. Foard, and L. B. Smith. SR-47 (1976), 137.

Perception of Second-Formant Transitions in Non-Speech Patterns. E. F. Kirstein. SR-7/8 (1966), 9.1.

Perception of Semi-Vowels and Laterals. M. P. Haggard. SR-17/18 (1969), 143.

The Perception of Speech. C. J. Darwin. SR-42/43 (1975), 59.

The Perception of Speech. M. Studdert-Kennedy. SR-23 (1970), 15.

Perception of Speech and Nonspeech, with and without Transitions. J. E. Cutting. SR-33 (1973), 37.

Perception of Speech and Nonspeech, with Speech-Relevant and Speech-Irrelevant Transitions. J. E. Cutting. SR-35/36 (1973), 55.

Perception of the Speech Code. A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. SR-9 (1967), 1.1.

Perception of Voiced and Voiceless Stop Consonants in Three-Year-Old Children. J. M. S. Winterkorn, P. F. MacNeilage, and M. S. Preston. SR-11 (1967), 41.

The Perception of Vowel Duration in VC and CVC Syllables. L. J. Raphael, M. F. Dorman, and A. M. Liberman. SR-42/43 (1975), 277.

Perceptual Competition Between Speech and Nonspeech. R. S. Day and J. E. Cutting. SR-24 (1970), 35.

Perceptual Integration and Selective Attention in Speech Perception: Further Experiments on Intervocalic Stop Consonants. B. H. Repp. SR-49 (1977), 37.

Perceptual Processing Time for Consonants and Vowels. D. B. Pisoni. SR-31/32 (1972), 83.

A Perceptual Study of English Diphthongs. T. J. Gay. (Ph.D. dissertation) Supplement to the Status Report, June 1967.

On Peripheral and Central Processes in Vision: Inferences from an Information-Processing Analysis of Masking with Patterned Stimuli. M. T. Turvey. SR-28 (1971), 1.

Perspectives in Vision: Conception or Perception? M. T. Turvey. SR-42/43 (1975), 1.

The Phi Coefficient as an Index of Ear Differences in Dichotic Listening. G. M. Kuhn. SR-29/30 (1972), 75.

Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and the Chimpanzee. P. Lieberman, E. S. Crelin, and D. H. Klatt. SR-24 (1970), 57.

Phonetic Activity in Reading: An Experiment with Kanji. D. Erickson, I. G. Mattingly, and M. T. Turvey. SR-33 (1973), 137.

Phonetic Aspects of Time and Timing. L. Lisker. SR-47 (1976), 113.

Phonetic Coding of Kanji. D. Erickson, I. G. Mattingly, and M. Turvey. (Abstract) SR-29/30 (1972), 155.

Phonetic Coding of Words in a Taxonomic Classification Task. G. C. Ellison. SR-41 (1975), 95.

A Phonetic-Context Controlled Strategy for Segmentation and Phonetic Labeling of Speech. P. Mermelstein. SR-37/38 (1974), 191.

Phonetic Feature Analyzers and the Processing of Speech in Infants. J. E. Cutting and P. D. Eimas. SR-37/38 (1974), 45.

Phonetic Prerequisites for First-Language Acquisition. I. G. Mattingly. SR-34 (1973), 65.

Phonetic Recoding and Reading Difficulty in Beginning Readers. L. S. Mark, D. Shankweiler, I. Y. Liberman, and C. A. Fowler. SR-49 (1977), 71.

Phonetic Segmentation and Recoding in the Beginning Reader. I. Y. Liberman, D. Shankweiler, A. M. Liberman, C. Fowler, and F. W. Fischer. SR-39/40 (1974), 105.

Phonetics: An Overview. A. S. Abramson. SR-23 (1970), 3.

Phonological Fusion in Synthetic and Natural Speech. J. E. Cutting. SR-33 (1973), 19.

Phonological Fusion of Stimuli Produced by Different Vocal Tracts. J. E. Cutting. SR-34 (1973), 61.

Phonological Fusion of Synthetic Stimuli in Dichotic and Binaural Presentation Modes. J. E. Cutting. SR-34 (1973), 55.

On the Physical Bases of Phonetic Features. P. Lieberman. SR-15/16 (1968), 137.

Physiological Aspects of Articulatory Behavior. K. S. Harris. SR-23 (1970), 49.

Physiological Aspects of Certain Laryngeal Features in Stop Production. H. Hirose, L. Lisker, and A. S. Abramson. SR-31/32 (1972), 183.

Physiological Aspects of Speech Production. K. S. Harris. SR-48 (1976), 21.

The Physiological Control of Durational Differences between Vowels Preceding Voiced and Voiceless Consonants in English. L. J. Raphael. SR-39/40 (1974), 199.

Pitch Determination by Adaptive Autocorrelation Method. G. Lukatela. SR-33 (1973), 185.

Pitch in the Perception of Voicing States in Thai: Diachronic Implications. A. S. Abramson. SR-41 (1975), 165.

"Posner's Paradigm" and Categorical Perception: A Negative Study. B. H. Repp. SR-45/46 (1976), 153.

Predicting Initial Cluster Frequencies by Phonemic Difference. J. E. Cutting. SR-42/43 (1975), 233.

Preliminaries to a Theory of Action with Reference to Vision. M. T. Turvey. SR-41 (1975), 1.

A Preliminary Electromyographic Study of Labial and Laryngeal Muscles in Danish Stop Consonant Production. E. Fischer-Jørgensen and H. Hirose. SR-39/40 (1974), 231.

A Preliminary Report on Six Fusions in Auditory Research. J. E. Cutting. SR-31/32 (1972), 93.

Primate Vocalizations and Human Linguistic Ability. P. Lieberman. SR-13/14 (1968), 65.

Problems in Machine Conversion of Print to "Speech." J. H. Gaitenby. SR-5/6 (1966), 7.1.

Procedures for Examining the Perceptual Capabilities for Processing Speech Sounds in Children Under Two Years of Age. M. S. Preston, J. Phillips, and G. Yeni-Komshian. SR-13/14 (1968), 185.

Processing Two Dimensions of Nonspeech Stimuli: The Auditory-Phonetic Distinction Reconsidered. M. J. Blechner, R. S. Day, and J. E. Cutting. SR-42/43 (1975), 221.

Programming for the Glace-Holmes Synthesizer. I. G. Mattingly. SR-17/18 (1969), 9.

On Pushing the Voice-Onset-Time (VOT) Boundary About. L. Lisker, A. M. Liberman, D. Erickson, and D. Dechovitz. SR-42/43 (1975), 257.

In (Qualified) Defense of VOT. L. Lisker. SR-44 (1975), 109.

Reaction Times to Comparisons Within and Across Phonetic Categories: Evidence for Auditory and Phonetic Levels of Processing. D. B. Pisoni and J. Tash. SR-34 (1973), 77.

Reading Aids for the Blind: A Special Case of Machine-to-Man Communication. F. S. Cooper, J. H. Gaitenby, I. G. Mattingly and N. Umeda. SR-17/18 (1969), 117.

Reading and the Awareness of Linguistic Segments. I. Y. Liberman, D. Shankweiler, B. Carter, and F. W. Fischer. SR-31/32 (1972), 145.

Reading, the Linguistic Process, and Linguistic Awareness. I. G. Mattingly. SR-27 (1971), 23.

The Relationships Between Speech and Reading. I. G. Mattingly and J. F. Kavanagh. SR-29/30 (1972), 83.

On the Relationship of Speech to Language. J. E. Cutting and J. F. Kavanagh. SR-41 (1975), 59.

A Reply to "Identification and Discrimination Functions for a Visual Continuum and Their Relation to the Motor Theory of Speech Perception," by Cross, Lane, and Sheppard. A. M. Liberman, M. Studdert-Kennedy, K. S. Harris, and F. S. Cooper. SR-3 (1965), 3.1.

Results of a VCV Spectrogram-Reading Experiment. G. M. Kuhn and R. McI. McGuire. SR-39/40 (1974), 67.

Right-ear Advantage for Musical Stimuli Differing in Rise-Time. M. J. Blechner. SR-47 (1976), 63.

A Right-Ear Advantage in Choice Reaction Time to Monaurally Presented Vowels: A Pilot Study. M. Studdert-Kennedy. SR-31/32 (1972), 75.

A Right-Ear Advantage in the Retention of Words Presented Monaurally. M. T. Turvey, D. Pisoni, and J. F. Croog. SR-31/32 (1972), 67.

Rise Time in Nonlinguistic Sounds and Models of Speech Perception. J. E. Cutting, B. S. Rosner, and C. F. Foard. SR-41 (1975), 71.

The Role of Auditory Short-Term Memory in Vowel Perception. D. B. Pisoni. SR-34 (1973), 89.

The Role of the Extrinsic and Intrinsic Tongue Muscles in Differentiating the English Tense-Lax Vowel Pairs. L. J. Raphael and F. Bell-Berti. SR-33 (1973), 203.

On the Role of Formant Transitions in Vowel Recognition. B. E. F. Lindblom and M. Studdert-Kennedy. SR-9 (1967), 7.1.

The Role of Speech in Language: Introduction to the Conference. A. M. Liberman. SR-37/38 (1974), 1.

Rules for Word Stress Analysis for Conversion of Print to Synthetic Speech. J. H. Gaitenby. SR-12 (1967), 131.

The Scaling of Speech Sounds. M. D. Vinegrad. SR-7/8 (1966), 8.1.

The Scaling of Synthetic Speech. M. D. Vinegrad. SR-11 (1967), 37.

In Search of the Acoustic Cues. A. M. Liberman and F. S. Cooper. SR-19/20 (1969), 9.

Segmentation of the Spoken Word and Reading Acquisition. I. Y. Liberman. SR-33 (1973), 157.

Selective Listening for Dichotically Presented Stop Consonants and Vowels. E. F. Kirstein and D. P. Shankweiler. SR-17/18 (1969), 133.

Selective Listening for Temporally Staggered Dichotic CV Syllables. E. F. Kirstein. SR-21/22 (1970), 63.

Separate Speech and Nonspeech Processing in Dichotic Listening. R. S. Day and J. C. Bartlett. SR-28 (1971), 93.

Shifts in Vowel Perception as a Function of Speaking Rate. R. R. Verbrugge, D. Shankweiler, W. Strange, and T. R. Edman. SR-47 (1976), 165.

Short-Term Habituation of the Infant Auditory Evoked Response. M. F. Dorman and R. Hoffmann. SR-29/30 (1972), 121.

On the Short-Term Retention of Serial, Tactile Stimuli. E. V. Sullivan and M. T. Turvey. SR-33 (1973), 123.

A Simple Model of Response Selection in the Dichotic Two-Response Paradigm. B. H. Repp. SR-49 (1977), 187.

A Simple Program for Synthesizing British English. M. P. Haggard and I. G. Mattingly. SR-12 (1967), 117.

A Simple Program for Synthesizing Speech by Rules. M. P. Haggard. SR-9 (1967), 5.1.

Some Acoustic Measures of Anticipatory and Carryover Coarticulation. F. Bell-Berti and K. S. Harris. SR-42/43 (1975), 297.

Some Aspects of Coarticulation. F. Bell-Berti and K. S. Harris. SR-45/46 (1976), 197.

Some Aspects of the Production of Oral and Nasal Labial Stops. K. S. Harris, G. F. Lysaught, and M. H. Schvey. SR-2 (1965), 1.1.

Some Aspects of Selective Readout from Iconic Storage. M. T. Turvey. SR-29/30 (1972), 1.

Some Attempts to Make Vowels Behave Like Speech. T. G. Halwes. SR-11 (1967), 57.

Some Characteristics of Perception in the Speech Mode. A. M. Liberman. SR-15/16 (1968), 39.

Some Effects of Context on Voice Onset Time in English Stops. L. Lisker and A. S. Abramson. SR-5/6 (1966), 1.1.

Some Effects of Oral Anesthesia upon Speech: An Electromyographic Investigation. G. J. Borden. SR-29/30 (1972), 27.

Some Electromyographic Measures of Coarticulation in VCV Utterances. T. Gay. SR-44 (1975), 137.

Some Electromyographic Observations of the Speech of Persons with Impaired Hearing. D. A. Huntington and K. S. Harris. SR-3 (1965), 5.1.

Some Maskinglike Phenomena in Speech Perception. M. F. Dorman, L. J. Raphael, A. M. Liberman, and B. Repp. SR-42/43 (1975), 265.

Some Observations on the Efficiency of Speech Sounds. A. M. Liberman, F. S. Cooper, M. Studdert-Kennedy, K. S. Harris, and D. P. Shankweiler. SR-4 (1965), 1.1.

Some Observations on a Model for Speech Perception. A. M. Liberman, F. S. Cooper, K. S. Harris, P. F. MacNeilage, and M. Studdert-Kennedy. SR-1 (1965), 2.1.

Some Observations on Monosyllable Production by Deaf and Dysarthric Speakers. D. A. Huntington, K. S. Harris, D. P. Shankweiler, and G. N. Sholes. SR-10 (1967), 6.1.

Some Rules for Synthesis of General American English. I. G. Mattingly. SR-13/14 (1968), 37.

Some Stress Effects on Electromyographic Measures of Consonant Articulations. K. S. Harris, T. J. Gay, G. N. Sholes, and P. Lieberman. SR-13/14

(1968), 137.

The Specialization of the Language Hemisphere. A. M. Liberman. SR-31/32 (1972), 1.

Speech, the Alphabet and Teaching to Read. I. Y. Liberman and D. Shankweiler. SR-48 (1976), 75.

The Speech Code and the Physiology of Language. I. G. Mattingly and A. M. Liberman. SR-15/16 (1968), 9.

Speech Cues and Sign Stimuli. I. G. Mattingly. SR-27 (1971), 89.

Speech Misperception: Inferences About a Cue for Cluster Perception from a Phonological Fusion Task. J. E. Cutting. SR-33 (1973), 57.

On the Speech of Neanderthal Man. P. Lieberman and E. S. Crelin. SR-21/22 (1970), 157.

Speech and the Problem of Perceptual Constancy. D. Shankweiler, W. Strange, and R. Verbrugge. SR-42/43 (1975), 117.

Speech Perception. M. Studdert-Kennedy. SR-39/40 (1974), 1.

A Speech Perception Paradox?: The Right-Ear Advantage and the Lag Effect. R. A. Weeks. SR-33 (1973), 29.

Speech Production and Perception in a Patient with Severe Impairment of Somesthetic Perception and Motor Control. P. F. MacNeilage, T. P. Rootes, and R. A. Chase. SR-5/6 (1966), 5.1.

Speech Recognition Through Spectrogram Matching. F. Ingemann and P. Mermelstein. SR-39/40 (1974), 53.

Speech Synthesis for Phonetic and Phonological Models. I. G. Mattingly. SR-23 (1970), 117.

Speech Synthesis by Rule as a Research Technique. I. G. Mattingly. SR-10 (1967), 3.1.

State-of-the-Art Report on Language Processing. A. M. Liberman. SR-35/36 (1973), 107.

Static and Dynamic Acoustic Cues in Distinctive Tones. A. S. Abramson. SR-47 (1976), 121.

Stimuli for Dichotic Listening Experiments. C. J. Darwin. SR-11 (1967), 61.

Stimulus Versus Response Competition and Ear Asymmetry in Dichotic Listening. C. J. Darwin. SR-25/26 (1971), 29.

Stop Consonant Recognition: Release Bursts and Formant Transitions as Functionally Equivalent, Context-Dependent Cues. M. Dorman, M. Studdert-

Kennedy, and L. J. Raphael. SR-47 (1976), 1.

Stop Consonant Voicing and Pharyngeal Cavity Size. F. Bell-Berti and H. Hirose. SR-31/32 (1972), 207.

Stop Duration and Voicing in English. L. Lisker. SR-19/20 (1969), 27.

Stop Voicing Production: Natural Outputs and Synthesized Inputs. L. Lisker. SR-47 (1976), 155.

Stress and the Elastic Syllable: An Acoustic Method for Delineating Lexical Stress Patterns in Connected Speech. J. H. Gaitenby. SR-41 (1975), 137.

Stress and Syllable Duration Change. K. S. Harris. SR-35/36 (1973), 31.

Studies on the Development of Stop Consonants in Children. M. S. Preston and G. Yeni-Komshian. SR-11 (1967), 49.

Study of Non-Speech Voluntary Palate Movements by Scaling and Electromyographic Techniques. R. L. Shelton, K. S. Harris, G. N. Sholes, and P. M. Dooley. SR-12 (1967), 79.

A Study of Prosodic Features. P. Lieberman. SR-23 (1970), 179.

A Study of the Relationship Between Production and Perception of Initial and Intervocalic /t/ and /d/ in Individual English-Speaking Adults. B. Lorge. SR-9 (1967), 3.1.

A Study of Semantic and Psychophysical Test Responses to Controlled Variations in Fundamental Frequency. K. Hadding-Koch and M. Studdert-Kennedy. SR-1 (1965), 4.1.

A Study of Voicing in Initial Stops Found in the Pre-Linguistic Vocalizations of Infants from Different Language Environments. M. S. Preston, G. Yeni-Komshian, and R. E. Stark. SR-10 (1967), 7.1.

A Study of Voicing in the Stops Found in the Pre-Linguistic Babblings of Infants. M. S. Preston. SR-7/8 (1966), 6.1.

The Stuttering Larynx: An EMG, Fiberoptic Study of Laryngeal Activity Accompanying the Moment of Stuttering. F. J. Freeman and T. Ushijima. SR-41 (1975), 217.

Supraglottal Air Pressure in the Production of English Stops. L. Lisker. SR-4 (1965), 3.1.

Syllable Synthesis. I. G. Mattingly. SR-49 (1977), 111.

Synthesis by Rule of General American English. I. G. Mattingly. SR-7/8 (1966), 4.1.

Synthesis by Rule of General American English. I. G. Mattingly. (Ph.D dissertation) Supplement to the Status Report, April 1968.

Synthesis by Rule as a Tool for Phonological Research. I. G. Mattingly. SR-19/20 (1969), 61.

Synthetic Speech Comprehension: A Comparison of Listener Performances with and Preferences Among Different Speech Forms. P. W. Nye, F. Ingemann, and L. Donald. SR-41 (1975), 117.

A Systems Approach to the Cerebral Hemispheres. C. A. Fowler. SR-44 (1975), 17.

Temporal Factors in Perception of Dichotically Presented Stop Consonants and Vowels. E. F. Kirstein. (Ph.D. dissertation) SR-24 (1970), Part II, 1.

Temporal Order Judgments in Speech: Are Individuals Language-Bound or Stimulus-Bound? R. S. Day. SR-21/22 (1970), 71.

Temporal Order Perception of a Reversible Phoneme Cluster. R. S. Day. SR-24 (1970), 47.

Testing Synthesis-by-Rule with the OVEBORD Program. F. Ingemann. SR-41 (1975), 127.

Thai Tones as a Reference System. A. S. Abramson. SR-44 (1975), 127.

On Time and Timing in Speech. L. Lisker. SR-23 (1970), 151.

Tonal Experiments in Whispered Thai. A. S. Abramson. SR-19/20 (1969), 37.

The Tones of Central Thai: Some Perceptual Experiments. A. S. Abramson. SR-39/40 (1974), 91.

Towards a Unified Phonetic Theory. P. Lieberman. SR-17/18 (1969), 157.

Transillumination of the Larynx in Running Speech. L. Lisker, A. S. Abramson, F. S. Cooper, and M. H. Schvey. SR-5/6 (1966), 4.1.

A Transistorized Formant-Type Synthesizer. R. Epstein. SR-1 (1965), 7.1.

Two Processes in Vowel Recognition: Inferences from Studies of Backward Masking. M. F. Dorman, D. Kewley-Port, S. Brady, and M. T. Turvey. SR-37/38 (1974), 233.

Two Questions in Dichotic Listening. M. Studdert-Kennedy. SR-41 (1975), 51.

Universals in Phonetic Structure and Their Role in Linguistic Communication. M. Studdert-Kennedy. SR-48 (1976), 43.

Velar Activity in Voicing Distinctions: A Simultaneous Fiberoptic and Electromyographic Study. F. Bell-Berti and H. Hirose. SR-31/32 (1972), 223.

Velar Movement and Its Motor Command. T. Ushijima and H. Hirose. SR-41 (1975), 207.

Velopharyngeal Function in Oral/Nasal Articulation and Voicing Gestures. F. B. Berti and H. Hirose. SR-28 (1971), 143.

The Velopharyngeal Mechanism: An Electromyographic Study. F. Bell-Berti. (Ph.D. dissertation). Supplement to the Status Report, September 1973.

The Velopharyngeal Mechanism: An Electromyographic Study. A Preliminary Report. F. B. Berti. SR-25/26 (1971), 117.

Visual Processing and Short-Term Memory. M. Turvey. SR-48 (1976), 97.

Visual Storage or Visual Masking?: An Analysis of the "Retroactive Contour Enhancement" Effect. M. T. Turvey, C. F. Michaels, and D. K. Port. SR-31/32 (1972), 131.

Vocal Tract Limitations on the Vowel Repertoires of Rhesus Monkey and Other Nonhuman Primates. P. Lieberman, D. H. Klatt, and W. A. Wilson. SR-15/16 (1968), 157.

Vocal Tract Normalization for /s/ and /š/. J. May. SR-48 (1976), 67.

Vocal Tract Size Normalization in the Perception of Stop Consonants. T. C. Rand. SR-25/26 (1971), 141.

A Voice for the Laboratory Computer. F. S. Cooper, T. C. Rand, R. S. Music, and I. G. Mattingly. SR-24 (1970), 91.

Voice Onset Time in the Production and Perception of English Stops. L. Lisker and A. S. Abramson. SR-1 (1965), 3.1.

Voice Onset Time in Stop Consonants: Acoustic Analysis and Synthesis. A. S. Abramson and L. Lisker. SR-3 (1965), 1.1.

Voice Timing: Cross-Language Experiments in Identification and Discrimination. A. S. Abramson and L. Lisker. SR-13/14 (1968), 49.

Voice Timing in Korean Stops. A. S. Abramson and L. Lisker. SR-27 (1971), 179.

Voice-Timing Perception in Spanish Word-Initial Stops. A. S. Abramson and L. Lisker. SR-29/30 (1972), 15.

The Voicing Dimension: Some Experiments in Comparative Phonetics. L. Lisker and A. S. Abramson. SR-11 (1967), 9.

Vowel Duration as a Cue to the Perceptual Separation of Cognate Sounds in American English. L. J. Raphael. (Ph.D. dissertation) Abstract: SR-25/26 (1971), 147. Text: Supplement to the Status Report, August 1971.

Vowel and Nasal Durations as Perceptual Cues to Voicing in Word-Final Stop Consonants. M. F. Dorman, L. J. Raphael, F. Freeman, and C. Tobin. SR-37/38 (1974), 263.

Vowel and Nasal Durations in Vowel-Nasal-Consonant Sequences in American English: Spectrographic Studies. L. J. Raphael, M. F. Dorman, C. Tobin, and F. Freeman. SR-37/38 (1974), 255.

Vowel Stress and Articulatory Reorganization. K. S. Harris. SR-28 (1971), 167.

Weak Syllables in a Primitive Reading-Machine Algorithm. G. Sholes. SR-45/46 (1976), 163.

What Can /w/, /l/, /y/ Tell Us About Categorical Perception? L. Frazier. SR-48 (1976), 235.

What Information Enables a Listener to Map a Talker's Vowel Space? R. Verbrugge, W. Strange, and D. Shankweiler. SR-37/38 (1974), 199.

What Information Enables a Listener to Map a Talker's Vowel Space? R. R. Verbrugge, W. Strange, D. P. Shankweiler, and T. R. Edman. SR-45/46 (1976), 63.

Word and Phrase Stress by Rule for a Reading Machine. J. H. Gaitenby, G. N. Sholes, and G. M. Kuhn. SR-29/30 (1972), 105.

Word-Final Stops in Thai. A. S. Abramson. SR-28 (1971), 193.

Word Recall in Aphasia. D. Kewley-Port. SR-39/40 (1974), 123.

Why are Spectrograms Hard to Read? A. M. Liberman, F. S. Cooper, and M. Studdert-Kennedy. SR-10 (1967), 1.1.

III. PUBLICATIONS AND REPORTS

IV. APPENDIX

PUBLICATIONS AND REPORTS

Blechner, Mark J. (1977) Left-ear advantage for sounds characterized by a rapidly varying resonance frequency. *Bull. Psychonom. Soc.*, vol. 9(5), 363-366.

Borden, Gloria J., M. F. Dorman, F. J. Freeman and L. J. Raphael. (1977) Electromyographic changes with delayed auditory feedback of speech. *J. Phonetics*., *5*, 1-8.

Erickson, Donna M., I. G. Mattingly and M. T. Turvey. (in press) Phonetic activity in reading: an experiment with Kanji. *Lang. Sp*.

__________, H. L. Fitch, T. G. Halwes, A. M. Liberman. (1977) Trading relation in perception between silence and spectrum. *J. Acoust. Soc. Am*., vol. 61, Suppl. no. 1, (S)46.

Jusczyk, Peter W., B. S. Rosner, J. E. Cutting, C. L. Foard and L. B. Smith. (1977) Categorical perception of nonspeech sounds by 2-month-old infants. *Percept. Psychophys*., vol. 21(1), 50-54.

Repp, Bruno H. (1977) Further observations on the function relating the voicing boundary to changes along a place continuum. *J. Acoust. Soc. Am*., *61*, Suppl. No. 1, (A)S47.

Shankweiler, D., W. Strange and R. R. Verbrugge. (1977) Speech and the problem of perceptual constancy. In *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, ed. by R. E. Shaw and J. D. Bransford. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Strange, W., R. R. Verbrugge, D. P. Shankweiler and T. R. Edman. (1976) Consonant environment specifies vowel identity. *J. Acoust. Soc. Am*., *60*, 213-224.

Verbrugge, Robert R. (1977) Resemblances in language and perception. In *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, ed. by R. E. Shaw and J. D. Bransford. (Hillsdale, N.J.: Erlbaum Associates).

__________ and D. P. Shankweiler. (1977) Prosodic information for vowel identity. *J. Acoust. Soc. Am*., *61*, Suppl. No. 1, S39(A).

__________, D. P. Shankweiler, W. Strange and T. R. Edman. (1976) Shifts in vowel perception as a function of speaking rate. *J. Acoust. Soc. Am*., *59* No. 1, S5(A).

__________, W. Strange, D. P. Shankweiler and T. R. Edman. (1976) What information allows a listener to map a talker's vowel space? *J. Acoust. Soc. Am*., *60*, 198-212.

APPENDIX

DDC (Defense Documentation Center) and ERIC (Educational Resources Information Center) numbers SR-21/22 to SR-48:*

Status Report		DDC*	ERIC*
SR-21/22	January - June 1970	AD 719382	ED-044-679
SR-23	July - September 1970	AD 723586	ED-052-654
SR-24	October - December 1970	AD 727616	ED-052-653
SR-25/26	January - June 1971	AD 730013	ED-056-560
SR-27	July - September 1971	AD 749339	ED-071-533
SR-28	October - December 1971	AD 742140	ED-061-837
SR-29/30	January - June 1972	AD 750001	ED-071-484
SR-31/32	July - December 1972	AD 757954	ED-077-285
SR-33	January - March 1973	AD 762373	ED-081-263
SR-34	April - June 1973	AD 766178	ED-081-295
SR-35/36	July - December 1973	AD 774799	ED-094-444
SR-37/38	January - June 1974	AD 783548	ED-094-445
SR-39/40	July - December 1974	AD A007342	ED-102-633
SR-41	January - March 1975	AD A103325	ED-109-722
SR-42/43	April - September 1975	AD A018369	ED-117-770
SR-44	October - December 1975	AD A023059	ED-119-273
SR-45/46	January - June 1976	AD A026196	ED-123-678
SR-47	July - September 1976	AD A031789	ED-128-870
SR-48	October - December 1976	AD A036735	ED-135-028
SR-49	January - March 1977	**	**
SR-50	April - June 1977	**	**

*See next page for ordering information.

**DDC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Behavior Abstracts*, P.O. Box 22206, San Diego, California 92122.